The Dissertation Committee for Erik Torres Hernandez Certifies that this is the approved version of the following dissertation:

Lessons Learned from Studying Peptide Chemistry and the Development in Mimicking Its Modularity in the Design of a New Oligomer Using Guanidiniums, Alpha-Diketones, and Boronic Acids

Committee:

Eric V. Anslyn, Supervisor

Edward M. Marcotte

Kevin Dalby

Adrian Keatinge-Clay

Emily L. Que

Lessons Learned from Studying Peptide Chemistry and the Development in Mimicking Its Modularity in the Design of a New Oligomer using Guanidiniums, Alpha-Diketones, and Boronic Acids

By

Erik Torres Hernandez, B. A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Dedication

A dedication is only an opportunity to mention someone no one else cares to know. Therefore, I dedicate this dissertation to something more general: the student collective for all its efforts made to make our principal investigators look good.

Acknowledgements

I would like to acknowledge persistence. Without it, no one ever finishes something like graduate school.

Lessons Learned from Studying Peptide Chemistry and the Development in Mimicking Its Modularity in the Design of a New Oligomer using Guanidiniums, Alpha-Diketones, and Boronic Acids

Erik Torres Hernandez, Ph. D.

The University of Texas at Austin, 2017

Supervisor: Eric V. Anslyn

Peptide chemistry is a versatile tool in the development of diverse compounds that can be applied in various contexts: biology, drug development, organic synthesis, chemo sensing, combinatorial library design, and unnatural foldamer formation. This chemistry is highly optimized that commercial instruments are available for the fast production of these compounds. Despite the optimization, there are new, engaging avenues scientists can still explore in this well-established field. Our work in the Anslyn group is currently focused to incorporate more functionality into peptides beyond the repertoire given by nature. These efforts are targeted for many uses as described in this dissertation. The lessons learned from synthetic peptide design have directed our group also to explore the development of new oligomer systems that take a modular approach in its elongation as used in peptide chemistry. The use of synthetically accessible monomers that are linked by a reliable reaction is the simple, but powerful guiding principle behind the work presented here.

Chapter 2 describes the development of a methodology to modify peptides in a sequential and selective manner targeting the most reactive amino acids. The chemistry presented in this chapter serves as the basis for the synthesis of model peptides with

fluorescent probes. In chapter 3, these fluorescently labeled peptides are used for proof-of-concept studies for a new single-molecule protein-sequencing platform. Inspiration taken from chapters 2 and 3 brings chemistries facilitating the introduction of ortho-(aminomethyl)boronic acids into peptides. Chapter 5 explores chemical modifications ligands allowing for induced higher-order structuring. Finally, Chapter 6 leaves the peptide space to explore a supramolecular oligomer that will serve as the foundation for a new backbone branched with side chain functionalities.

Table of Contents

List of Tables xv

List of Figures xvi

Chapter 1: Peptide and Unnatural Oligomer Chemistries.....1

1.1 Introduction.....1

1.2 Abridged Literature Review 1

1.2.1 Solution-Phase Peptide Synthesis: Short Peptides Initially Made 3

1.2.2 Solid-Phase Peptide Synthesis: Preferred mdfadsethod for Short Polyamides 5

1.2.3 Natural Residue Synthetic Peptides: Chemistry Approaches to Making Peptide 9

1.2.4 Unnatural Synthetic Peptdies: Expanding the Residue Repertoire. 11

1.2.5 Foldamers: Expanding the Backbone Repertoire for Biomimicry 11

1.3 Summary 12

1.4 Dissertation Outlook 13

1.4.1 Chapter 2: Natural Side Chain Modifications.....13

1.4.2 Chapter 3: Fluorescent Peptides Used for a Single-Molecule Peptide Sequencing Platform.....14

1.4.3 Chapter 4: Secondary Amine Resude Peptides Alkylated with Boronic Acids 14

1.4.4 Chapter 5: Secondary Amine Residue Peptides Alkylated with Bipyridines to Induce Folding 14

1.4.5 Chapter 6: Modular Oligomer Design Achieved by Using Guanidinium, Boronic Acid, and Dione 15

1.5 References.....15

Chapter 2: Solution-Phase and Solid-Phase Sequential, Selective Modification of Side Chains in KDYWEC and KDYWE as Moedls for Usage in Single-Molecule Protein Sequencing 17

2.1 Introduction 17

2.1.1 Need for Selective Labeling of Amino Acid Residues for Emerging Single-Molecule Sequencing Technologies 17

2.1.2 Chemistries Available for Selective Modification of Peptides ...18

2.1.3 Scientific Aim: Modify Most Reactive Amino Acids on Model Peptides 18

2.2 Results and Discussion 21

2.2.1 Solution-Phase Sequential, Selective Modification of KDYWEC 22

2.2.2 Solid-Phase Sequential, Selective Modification of KDYWE 25

2.3 Conclusions 28

2.4 Experimental 28

2.4.1 General Materials 28

2.4.2 General Instrumentation 29

2.4.3 General Solid-Phase Model Peptide Synthesis 29

2.4.4 Solution-Phase Modification Studies KDYWEC 30

2.4.5 Solid-Phase Modification Studies of KDYWE 33

2.5 Acknowledgments 35

2.6 Experimental Characterization 35

2.7 References 60

Chapter 3: Fluorescently Labeled Model Peptdes for a New Single-Molecule Peptide Sequencing Platform Using Edman Degradation and Total Internal Reflection Flourescence Microsocopy 62

3.1 Introduction 62

3.1.1 Mass Spectrometry: Current Peptide Sequencing Approach and Limitations 62

3.1.2 Higher Sensitivity Required 63

3.1.3 Single-Molecule DNA Sequencing 63

3.1.4 Marcotte Single-Molecule Protein Sequencing (Marcotte Sequencing) 64

3.1.5 Theoretical Justification for Marcotte Sequencing 67
3.1.6 Competing Single-Molecule Protein Sequencing Approaches ...68
3.1.7 Model Peptides Required for Sequencing.....69
3.2 Results and Discussion 70
3.2.1 Description of Instrumentation and Data Analysis 70
3.2.2 Bulk Fluorescence Studies 72
3.2.3 N-Terminal Protecting Group and Dye Selection for Bulk Fluorescence Studies.....72
3.2.4 Bead Assay Results 74
3.2.5 Solution-Phase Confirmation of Edman Degradation with Fluorescent Side Chains.....74
3.2.6 Missellaneous fluorescent Peptides 75
3.2.7 Smaller Scale Reactions.....76
3.2.8 Atto 647 N Model Peptides Single-Molecule Sequencing: Single-Dye 76
3.2.9 Atto 647 N Model Peptides Single-Molecule Sequencing: Two Dyes 78
3.2.10 Future Sequencing Studies: Spacing Limitation for Fluorescently Labeled Side Chains 80
3.2.11 Future Sequencing Studies: Naturally Occuring Peptides 80
3.2.12 Future Sequencing Studies: Synthetic Insulin Digests 80
3.2.13 Labeling of Cystein Residues in Recombinant Human Insulin 81
3.2.14 Digestion of TMR Labeled Insulin Fragments 83
3.2.15 Silicone-Based Rhdoamine for N-Terminal Modification (Si-Rh) 85
3.3 Conclusion…….86
3.4 Experimental 86
3.4.1 General Materials 86
3.4.2 General Instrumentation.....87
3.4.3 Genral Solid-Phase Modl Peptide Synthesis.....88
3.4.4 Labeling of N-terminal ivDde/Fmoc-Potected Peptides with Cysteine or Lysine Residues 88

3.4.5 Synthesis N-Terminal boc-Protected Peptides with Cysteine or Lysine Residues 89
3.4.6 Solid-Phase N-Terminal Labeling of Peptides with NHS-Si-Rh 89
3.4.7 Acetylation of Peptides 89
3.4.8 Labeling of Human Recombinant Insulin 90
3.4.9 Glu-C Digestion TMR Labeled Chain and Chain B 90
3.5 Acknowledgements 90
3.6 Experimental Characterization 92
3.7 References 192

Chapter 4: An Efficient Methodology to Introduce o-(aminomethyl)phenyl-boronic Acids into Peptides: Alkylation of Secondary Amines 194
4.1 Introduction 194
4.1.1 Need for Facile Incorporation of o-(aminomethyl)phenyl-boronic Acids into Peptides: Alkylation of Secondary Amines 194
4.1.2 Current Approaches to Incorporate Boronic Acids into Peptides 194
4.1.3 Limitations with Current Approaches 195
4.1.4 Scientfic Aim: Use Alkylation Strategies and Secondary Amines to Incorporate *o*-(aminomethyl)phenyl-boronic Acids 196
4.2 Results and Discussion 197
4.2.1 Solution-Phase Incorporation of Secondary Amines and Alkyltion 197
4.2.2 Solid-Phase Incorporation of Secondary Amines and Alkylation 199
4.2.3 Building-Block Incorporation of Secondary Amines and Alkylation 200
4.3 Conclusions 202
4.4 Experimental 203
4.4.1 General Materials 203
4.4.2 General Instrumentation 204
4.4.3 General Procedure (A): Synthesis of Dimeric Peptides and N_{ε}-Methyl Lysine Peptides 204

4.4.4 General Procedure (B): Solution-Phase Alkylation with o-(bromomehyl)phenylboronic Acid.....205

4.4.5 General Procedure (C): Solid-Phase Modification of the N-Terminus with EZ-Link™ NHS-PEG_4-Biotin: Post-Automated Solid-Phase Synthesis 206

4.4.6 Synthesis of Boronic Acid Peptides.....206

4.5 Acknowledgements.....214

4.6 Experimental Characterization.....215

4.7 References.....251

Chapter 5: The Synthesis of Bipyridine Peptides via Alkylation with 5- and 6-(bromomethyl)-2,2'-bipyridine for Metal-Based Cyclization Studies in the Presence of Hydrazone Forming Peptides.....253

5.1 Inroduction.....253

5.1.1 Metals and Ligands in Forming Higher-Order Organic Structures 253

5.1.2 Facilitating Incorporation of Ligands into Peptides.....253

5.1.3 Bipyridine Ligands.....254

5.1.4 Relevance for Unnatural Metal-Based Peptides 255

5.1.5 The Synthesis of Metal-Binding Peptides.....255

5.1.6 Rationale for Alkylating Peptide with Bipyridine 257

5.1.7 Introduction of Dynamic Reversible Covalent Bonds 259

5.1.8 Induce Cyclization of Bipyridine Peptides in the Presence of Aldehyde and Hydrazide Peptides and the Characterization of Higher-Ordered Structgures.....260

5.2 Results and Discussion 261

5.2.1 Rationale for MP-6 and MP-5.....261

5.2.2 Methods to Characterize Cyclization.....262

5.2.3 Titration Studies with Small Molecule Bipyridines.....262

5.2.4 MP-6 and MP-5 Binding with Fe^{2+} and Zn^{2+} 265

5.2.5 Ensuring Cyclization of MP-5 267

5.2.6 Alternate Structure with Free Bipy: Oligomer.....268

5.2.7 MP Control Study 269

5.2.8 Alternative Small-Molecule Bipyridines and MP-5 269

5.2.9 Condensing HydPep-1 on Compound 5.5 and Complexation with MP-5 270
5.2.10 Condensing HydPep-2 on AldPep Followed by Condensing on Compound 5.5 272
5.2.10 Comlexation of Compound 5.7 273
5.3 Conclusions 274
5.4 Experimental 275
5.4.1 General Materials 275
5.4.2 General Instrumentation 275
5.4.3 General Procedure for Peptide Synthesis of MP-5 and MP-6 ..276
5.4.4 AldPep, HydPep-1, and HydPep-2 276
5.4.5 General Procedure for Alkylation of Peptides 277
5.4.6 Synthesis Brominated Ligand Precursors 277
5.4.7 Well-Plate Titration Experiments 277
5.4.8 UV-Vis Titration for Hydrazone Peptide and Mp-5 Mixtures..278
5.4.9 NMR Titration Experiments 278
5.5 Acknowledgements 278
5.6 Experimental Characterization 279
5.7 References 293

Chapter 6: Supramolecular Oligomer Formation of 1-([2,2':6',2"-terpyridin]-4'-ylmethyl)guanidine with Zn^{2+}, Thiophene-2,5-diyldiboronic Acid, and 2,3-Butadione 295
6.1 Inroduction 295
6.1.1 Unnatural, Synthetic Oligomer Design from a Peptide Chemist's Perspective 295
6.1.2 Limitations of Unnatural, Oligomer and Synthesis 296
6.1.3 Lessons from Peptide Chemistry that can be Applied to Unnatural Oligomers 298
6.1.4 Modular Oligomer Design 299
6.1.5 A Guanidinium-Terpyridine Supramolecular Oligomer 300
6.1.7 Possible Improvements When Compared to Peptide Chemistry 304

6.1.8 Dynamic Covalent Bonds to Overcome Issues with Sequence Reproducibility 305
6.1.9 The Dynamic/Orthogonal Nature of the Guanidinium-Dione-Boronic Acid Interaction 306
6.1.10 Scientific Aim: Order of Addition and Oligomer Formation Studies 307
6.2 Results and Discussion 308
6.2.1 Synthesis of Compound 6.1 308
6.2.2 Order of Addition Studies 309
6.2.3 Boron NMR Titration Studies 314
6.2.4 Proton NMR of Oligomerzation with Thiophene-2,5-diyldiboronic Acid (6.13) 318
6.2.5 DOSY NMR Studies 319
6.3 Conclusions 320
6.4 Experimental 321
6.4.1 General Material 349
6.4.2 General Instrumentation .....nn
6.4.3 Synthesis of 1-([2,2':6',2"-terpyridin]-4'-ylmethyl)guanidine (6.1) 321
6.4.4 Order of Addition Proton NMR Studies 323
6.4.5 Boron NMR Titration Studies 323
6.4.6 Proton DOSY NMR Experiments 324
6.5 Acknowledgements 324
6.6 Experimental Characterization 329
6.7 References 346
Bibliography 349

List of Tables

Table 4.1: Biotinylated peptide library sequences and yields. PEG_4-Biotin was placed on the N-terminus. X_2 = -$CH_2CH_2CH_2CH_2(NH(CH_3)(CH_2C_6H_6B(OH)_2))$ 203

List of Figures

Figure 1.1: Figure 1.1: Coupling Methodology Used for Synthesis of Benzoylpentaglycine. 3

Figure 1.2: Amide Coupling Reactants 4

Figure 1.3: Amide Coupling Reactants Used in Solid-Phase Peptide Synthesis (SPPS) 7

Figure 1.4: Cationic Species Formed from Deprotecting Side Chains. 8

Figure 1.5: :Native Chemical Ligation (NCL) 10

Figure 2.1: Selective Modification Route in Solution. 19

Figure 2.2: Solid-Phase Modification of KDYWE. 20

Figure 2.3: Peptide Intermediates Synthesized. 22

Figure 2.4: Solid-phase intermediates 35

Figure 2.5: ^{1}Proton NMR of 5,5-dimethyl-2-(3-methylbutanoyl)-3-oxocyclohex-1-en-1-yl Diethyl Phosphate (Phos-DOD). 37

Figure 2.6: ^{13}Carbon NMR of phos-DOD. 38

Figure 2.7: COSY NMR of phos-DOD 39

Figure 2.8: 2-D NMR for phos-DOD. 40

Figure 2.9: High-Resolution Mass Spectrometry Data for phos-DOD 42

Figure 2.10: High-Resolution Mass Spectrometry Data for Peptide 3 43

Figure 2.11: HPLC trace for purified peptide 3 44

Figure 2.12: High-Resolution Mass Spectrometry Data for Peptide 4 46

Figure 2.13: HPLC Trace for Purified Peptide 4 47

Figure 2.14: High-Resolution Mass Spectrometry Data for Peptide 5 48

Figure 2.15: HPLC Trace for Purified Peptide 5 49

Figure 2.16: High-Resolution Mass Spectrometry Data for Peptide 6. 50

Figure 2.17: HPLC Trace for Purified Peptide 6. 50
Figure 2.18: High-Resolution Mass Spectrometry Data for Peptide 8. 51
Figure 2.19: HPLC trace for purified peptide 8 52
Figure 2.20: High-Resolution Mass Spectrometry Data for Peptide 9 53
Figure 2.21: HPLC Trace for Purified Peptide 9 54
Figure 2.22: High-Resolution Mass Spectrometry Data for Peptide 10 55
Figure 2.23: HPLC Trace for Purified Peptide 10 56
Figure 2.24: High-Resolution Mass Spectrometry Data for Peptide 11 57
Figure 2.25: HPLC trace for purified peptide 11 58
Figure 2.26: High-Resolution Mass Spectrometry Data for Peptide 12 59
Figure 2.27: HPLC Trace for Purified Peptide 12 60
Figure 3.1: Edman Degradation Cycle 65
Figure 3.2: Schematic for the Overall Goal for the Marcotte Sequencing Approach 66
Figure 3.3: Fluorescently Labeled Peptides on a Glass Surface 67
Figure 3.4: Ideal Coverage of the Human Proteome Under Different Cleavage Conditions, Peptide Immobilization Anchoring Points, and side chains targeted. 68
Figure 3.5: Rhodamine Dyes 70
Figure 3.6: Marcotte Sequencing Platform 71
Figure 3.7: Bead Peptide Sequencing Study 75
Figure 3.8: Single-Molecule Sequencing Results for Peptides 3.13 and 3.14 78
Figure 3.9: Single-Molecule Sequencing Results for Peptides 3.15 and 3.16 79
Figure 3.10: Insulin (ins) Fragments Post Glu-C Digestion 81
Figure 3.11: Insulin Chain A Labeled with TMR Isolated by Syringe Filter 82

Figure 3.12: Insulin Chain B Labeled with TMR Isolated by Syringe Filter.....83
Figure 3.13: LRMS Ins. Fragments 84
Figure 3.14: 3.14: Silicone-Based Rhodamine (Si-Rh) 85
Figure 3.15: HRMS Data for Peptide 3.1 92
Figure 3.16: Analytical Trace for Peptide 3.1 93
Figure 3.17: HRMS Data for Peptide 3.2 94
Figure 3.18: HRMS Data for Peptide 3.2 95
Figure 3.19: HRMS Data for Peptide 3.3 96
Figure 3.20: HRMS Data for Peptide 3.3 97
Figure 3.21: HRMS Data for Peptide 3.4 98
Figure 3.22: HRMS Data for Peptide 3.4 99
Figure 3.23: Analytical Trace for Peptide 3.4 100
Figure 3.24: HRMS Data for Peptide 3.5 101
Figure 3.25: HRMS Data for Peptide 3.5 102
Figure 3.26: HRMS Data for Peptide 3.6 103
Figure 3.27: HRMS Data for Peptide 3.7 104
Figure 3.28: Analytical Trace for Peptide 3.7 105
Figure 3.29: HRMS Data for Peptide 3.8 106
Figure 3.30: HRMS Data for Peptide 3.8 107
Figure 3.31: Analytical Trace for Peptide 3.8 108
Figure 3.32: HRMS Data for Peptide 3.9 109
Figure 3.33: HRMS Data for Peptide 3.10 110
Figure 3.34: Analytical Trace for Peptide 3.10 111
Figure 3.35: HRMS Data for Peptide 3.11 112
Figure 3.36: Analytical Trace for Peptide 3.11 113

Figure 3.37: HRMS Data for Peptide 3.12 114
Figure 3.38: HRMS Data for Peptide 3.12 115
Figure 3.39: Analytical Trace for Peptide 3.12 116
Figure 3.40: HRMS Data for Peptide 3.13 117
Figure 3.41: HRMS Data for Peptide 3.13 118
Figure 3.42: Analytical Trace for Peptide 3.13 119
Figure 3.43: HRMS Data for Peptide 3.14 120
Figure 3.44: HRMS Data for Peptide 3.14 121
Figure 3.45: Analytical Trace for Peptide 3.14 122
Figure 3.46: HRMS Data for Peptide 3.15 123
Figure 3.47: HRMS Data for Peptide 3.16 124
Figure 3.48: HRMS Data for Peptide 3.17 125
Figure 3.49: Analytical Trace for Peptide 3.17 126
Figure 3.50: HRMS Data for Peptide 3.18 127
Figure 3.51: HRMS Data for Peptide 3.19 128
Figure 3.52: HRMS Data for Peptide 3.20 129
Figure 3.53: Analytical Trace for Peptide 3.20 130
Figure 3.54: LRMS Data for Peptide 3.21 131
Figure 3.55: LRMS Data for Peptide 3.22 132
Figure 3.56: LRMS Data for Peptide 3.23 133
Figure 3.57: HRMS Data for Peptide 3.24 134
Figure 3.58: HRMS Data for Peptide 3.24 135
Figure 3.59: HRMS Data for Peptide 3.25 136
Figure 3.60: Analytical Trace for Peptide 3.25 137
Figure 3.61: HRMS Data for Peptide 3.26 138

Figure 3.62: HRMS Data for Peptide 3.27 138
Figure 3.63: Analytical Trace for Peptide 3.27 140
Figure 3.64: HRMS Data for Peptide 3.28 141
Figure 3.65: HRMS data for peptide 3.29 142
Figure 3.66: AnalyticalTrace for Peptide 3.29 143
Figure 3.67: HRMS Data for Peptide 3.30 144
Figure 3.68: HRMS Data for Peptide 3.31 145
Figure 3.69: Analytical Trace for Peptide 3.31 146
Figure 3.70: HRMS Data for Peptide 3.32 147
Figure 3.71: HRMS Data for Peptide 3.33 148
Figure 3.72: Analytical trace for peptide 3.33 149
Figure 3.73: HRMS Data for Peptide 3.34 150
Figure 3.74: Analytical Trace for Peptide 3.34 151
Figure 3.75: HRMS Data for Peptide 3.35 152
Figure 3.76: Analytical Trace for Peptide 3.35 153
Figure 3.77: HRMS Data for Peptide 3.36 155
Figure 3.78: Analytical Trace for Peptide 3.36 156
Figure 3.79: HRMS Data for Peptide 3.37 156
Figure 3.80: Analytical Trace for Peptide 3.37 157
Figure 3.81: HRMS Data for Peptide 3.38 158
Figure 3.82: Analytical Trace for Peptide 3.38 159
Figure 3.83: HRMS Data for Peptide 3.39 160
Figure 3.84: Analytical Trace for Peptide 3.39 161
Figure 3.85: HRMS Data for Peptide 3.40 162
Figure 3.86: Analytical Trace for Peptide 3.40 163

Figure 3.87: HRMS Data for Peptide 3.41 164
Figure 3.88: Analytical Trace for Peptide 3.41 165
Figure 3.89: HRMS data for peptide 3.42 166
Figure 3.90: Analytical Trace for Peptide 3.42 167
Figure 3.91: HRMS Data for Peptide 3.43 168
Figure 3.92: Analytical Trace for Peptide 3.43 169
Figure 3.93: HRMS Data for Peptide 3.44 170
Figure 3.94: Analytical Trace for Peptide 3.44 171
Figure 3.95: HRMS data for peptide 3.45 172
Figure 3.96: Analytical Trace for Peptide 3.45 173
Figure 3.97: HRMS Data for Peptide 3.46 174
Figure 3.98: Analytical Trace for Peptide 3.46 175
Figure 3.99: HRMS Data for Peptide 3.47 176
Figure 3.100: Analytical Trace for Peptide 3.47.....177
Figure 3.101: HRMS Data for Peptide 3.48 178
Figure 3.102: Analytical Trace for Peptide 3.48.....179
Figure 3.103: HRMS Data for Peptide 3.49 180
Figure 3.104: Analytical Trace for Peptide 3.49.....181
Figure 3.105: HRMS Data for Peptide 3.50 182
Figure 3.106: Analytical Trace for Peptide 3.51.....183
Figure 3.107: HRMS Data for Peptide 3.52 184
Figure 3.108: Analytical Trace for Peptide 3.52.....185
Figure 3.109: HRMS Data for Peptide 3.53 186
Figure 3.110: Analytical Trace for Peptide 3.53.....187
Figure 3.111: HRMS Data for Peptide 3.54 188

Figure 3.112: Analytical Trace for Peptide 3.54 189
Figure 3.113: HRMS Data for Peptide 3.55 190
Figure 3.114: Analytical Trace for Peptide 3.55 191
Figure 4.1: Boronic Acids 195
Figure 4.2: Biotinylated Boronic Acid Peptides 201
Figure 4.3: Proton NMR Spectra for Compound *tert*-butyl (2-(2-chloro-*N*-methylacetamido)ethyl)(methyl)carbamate 214
Figure 4.4: Proton NMR Spectra for Compound 4.1 215
Figure 4.5: Proton NMR Spectra for Compound 4.3 216
Figure 4.6: Carbon NMR Spectra for Compound 4.3 217
Figure 4.7: HRMS Data for Compound 4.3 218
Figure 4.8: Purity Check for Compound 4.3 219
Figure 4.9: Proton NMR Spectra for Compound 4.4 220
Figure 4.10: Carbon NMR Spectra for Compound 4.4 221
Figure 4.11: HRMS for compound 4.4 222
Figure 4.12: Purity Check for Compound 4.4 222
Figure 4.13: Proton NMR Spectra for Compound 4.5 223
Figure 4.14: HRMS Data for Compound 4.5 224
Figure 4.15: Purity Check for Compound 4.5 225
Figure 4.16: Proton NMR Spectra for Compound 4.6 226
Figure 4.17: HRMS Data for Compound 4.6 227
Figure 4.18: Purity Check for Compound 4.6 227
Figure 4.19: Proton NMR Spectra for Compound 4.7 228
Figure 4.20: Carbon NMR spectra for compound 4.7 229
Figure 4.21: C/H correlation NMR for compound 4.7 230

Figure 4.22: HRMS data for compound 4.7 231
Figure 4.23: Purity check for compound 4.7 232
Figure 4.24: Proton NMR spectra for compound 4.8 233
Figure 4.25: Carbon NMR spectra for compound 4.8 234
Figure 4.26: HRMS Data for Compound 4.8 235
Figure 4.27: Purity Check for Compound 4.8 235
Figure 4.28: Proton NMR Spectra for Compound 4.9 236
Figure 4.29: Carbon NMR Spectra for Compound 4.9 237
Figure 4.30: HRMS Data for Compound 4.9 238
Figure 4.31: Purity Check for Compound 4.9 239
Figure 4.32: Proton NMR Spectra for Compound 4.10 240
Figure 4.33: Carbon NMR Spectra for Compound 4.10 241
Figure 4.34: HRMS data for compound 4.10 242
Figure 4.35: Purity Check for Compound 4.10 243
Figure 4.36: Proton NMR Spectra for Compound 4.11 244
Figure 4.37: Carbon NMR Spectra for Compound 4.11 245
Figure 4.38: C/H correlation NMR for compound 4.11 246
Figure 4.39: HRMS data for compound 4.11 247
Figure 4.40: Purity Check for Compound 4.11 248
Figure 4.41: Proton NMR Spectra for Compound 4.12 249
Figure 4.42: HRMS for Compound 4.12 250
Figure 4.43: Purity Check for Compound 4.12 251
Figure 5.1: Brominated Bipyridines 254
Figure 5.2: Proposed N-Methyl Lysine Modification with Bipyridines 257
Figure 5.3: Metallo-Peptides (MPs) Used for Cyclization Studies 258

Figure 5.4: : Hydrazone Forming Peptides Used Bipyridine Peptide Studies 260
Figure 5.5: Small Molecule Bipyridines Used for Titration Studies 262
Figure 5.6: Binding Isotherm for 5.1 263
Figure 5.7: Binding Isotherm for 5.5 264
Figure 5.8: Proton NMR for Zn^{2+} Titrated to a Solution of 5.6 264
Figure 5.9: Proton NMR Titration of Fe^{2+} to MP-6 265
Figure 5.10: Proton NMR Titration of Fe^{2+} to MP-5 266
Figure 5.11: UV-Vis Well-Plate Titration of MP-6 to Fe^{2+} 267
Figure 5.12: Binding Isotherm for Fe^{2+} and Compound 5.3 Titrated with MP-5 268
Figure 5.13: Binding Isotherm for Fe^{2+} and Compound 5.5 Titrated with MP-5 269
Figure 5.14: Proton NMR titration of Fe^{2+} to MP-5 and compound 5.6 270
Figure 5.15: Binding isotherm for MP-5, cyclic HydPep-1, and Fe^{2+} 271
Figure 5.16: Representative Structure Formed Between MP-5 and Cyclic HydPep-1 around Fe^{2+} 271
Figure 5.17: Representative structure formed with HydPep-2/AldPep dimer, hydrazide Tyrosine, and compound 5.5 (compound 5.7) 272
Figure 5.18: Binding Isotherm for MP-5, Compound 5.7, and Fe^{2+} 273
Figure 5.19: Possible structure formed when MP-5 and compound 5.7 bind around an Fe^{2+} 274
Figure 5.20: HRMS for MP-5 279
Figure 5.21: Analytical HPLC Trace for MP-5 279
Figure 5.22: HRMS for MP-6 280
Figure 5.23: Analytical HPLC Trace for MP-6 281

Figure 5.24: Well-Plate Titration for 2-2' bipyridine and Fe^{2+} 281
Figure 5.25: Proton NMR Titration of 2-2' Bipyridine with Fet^{2+}.....282
Figure 5.26: Proton NMR Titration of [2,2'-bipyridine]-5,5'-dicarbaldehyde with Fe^{2+} 282
Figure 5.27: Proton NMR Titration of [2,2'-bipyridine]-5,5'-dicarbaldehyde with Zn^{2+} 283
Figure 5.28: Proton NMR Titration of [2,2'-bipyridine]-5,5'-dicarbaldehyde with Zn^{2+} 283
Figure 5.29: Proton NMR Titration of MP-6 with Fe^{2+} 284
Figure 5.30: Well-Plate Titration for MP-5, 2,2'-bipyridine, and Fe^{2+}.....285
Figure 5.31: Well-Plate Titration for MP-5, 2,2'-bipyridine, and Fe^{2+}.....286
Figure 5.32: Well-Plate Titration for MP-5, 2,2'-bipyridine, and Fe^{2+}.....287
Figure 5.33: Proton NMR titration for MP-5, 2,2'-bipyridine, and Fe^{2+} 288
Figure 5.34: Binding Plot at 530 nm for MP-5 (0.206 mM), [2,2'-bipyridine]-5,5'-dicarbaldehyde (0.195 mM), and Fe^{2+} (0,214 mM).....288
Figure 5.35: Proton NMR Titration for MP-5, 2,2'-bipyridine, and Fe^{2+}.....289
Figure 5.36: Well-plate Titration for MP-6, 2,2'-bipyridine, and Fe^{2+}.....290
Figure 5.37: UV-Vis Titration of MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+} ..... …….291
Figure 5.38: MALDI Spectrum for MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+} .. …….291
Figure 5.39: UV-Vis Titration of MP-5, AldPep/HydPep-2 Bipyridine, and Fe^{2+} ... …….292
Figure 5.40: MALDI Spectrum for MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+} .. ……..292

Figure 6.1: Proposed Oligomer Using Guanidinyl Terpyridine, Zn^{2+}, Di-boronic acid, and Butadione.) 301
Figure 6.2: Comparing Peptide Backbone to Proposed Oligomer Backbone 303
Figure 6.3: Components Used in Order of Addition Studies and Orders Studied 310
Figure 6.4: Proton NMR Studies for A, D, B, C Order of Addition 311
Figure 6.5: Proton NMR Studies for B, A, D, C Order of Addition 313
Figure 6.6: Proton NMR Studies for C, A, D, B Order of Addition 314
Figure 6.7: Boron NMR Titration of Butadione into a Solution of Metal-Bound 6.1 and 6.11 316
Figure 6.8: Reactions of Compounds 6.11 and 6.12 with Butadione 317
Figure 6.9: Boron NMR Titration of Butadione into a Solution of Metal-Bound 6.1 and 6.12 318
Figure 6.10: ^{1}H NMR Titration of Butadione into a Solution of 6.13 and Metal-Bound 6.1 319
Figure 6.11: DOSY Experiment for a Mixture of Compounds 6.1, 6.13, and Butadione in DMSO 320
Figure 6.12: Proton and Carbon for Compound 6.6 326
Figure 6.13: HRMS for Compound for 6.6 327
Figure 6.14: Proton NMR for 6.7 328
Figure 6.15: HRMS for Compound for 6.7 329
Figure 6.16: Proton and Carbon for Compound 6.9 330
Figure 6.17: HRMS for Compound for 6.9 331
Figure 6.18: Proton and carbon for Compound 6.1 332
Figure 6.19: HRMS for Compound for 6.1. 333

Figure 6.20: Proton NMR Comparing Compounds 6.1 (A), 6.11 (B), 6.2 (C), and Zn^{2+} (D) 334
Figure 6.21: Proton NMR Comparing Compounds 6.1 to B,A; C,A; D,A Order of Addition 334
Figure 6.22: Proton NMR Comparing Aromatic Region Between Fully/Partially Bound 6.1 335
Figure 6.23: Proton NMR Comparing Guanidinyl Methylene Between Fully/Partially Bound 6.1 335
Figure 6.24: Proton NMR Comparing Compound 6.11 to a Mixture of 6.11 and 6.1 336
Figure 6.25: Proton NMR comparing orders of addition A, D, B; B, A, D; C, A, D for compounds 6.1, 6.11, and Zn^{2+} 336
Figure 6.26: Proton NMR comparing orders of addition C, A; C, A, D for compounds 6.1, 6.2, and Zn^{2+} 337
Figure 6.27: Proton NMR comparing compound 6.11 in the presence of Ac-Arg-OH 337
Figure 6.28: Proton NMR Comparing All Four Orders of Addition.....338
Figure 6.29: Proton NMR Comparing Compound 6.11 with All Four Orders of Addition 338
Figure 6.30: Proton NMR Comparing Compound 6.11 with All Orders of Addition and Different Conditions Applied to Increasing Signal Broadening Ddescribed in Section 6.2.2 339
Figure 6.31: Proton NMR Comparing Compounds 6.9 and 6.11 Individually and Mixtures 340

Figure 6.32: Proton NMR Comparing Benzyl Guanidinium and Compound 6.11 340
Figure 6.33: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.13, and Zn^{2+} for Methylene Group of 6.1 341
Figure 6.34: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.13, and Zn^{2+} 341
Figure 6.35: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.13, and Zn^{2+} for Methylene Group of 6.1 342
Figure 6.36: Comparing compound 6.1 and 6.14 342
Figure 6.37: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.14, and Zn^{2+} 343
Figure 6.38: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.14, and Zn^{2+} 343
Figure 6.39: Proton NMR Study of Titrating 6.2 to a Solution of 6.1, 6.14, and Zn^{2+} 344
Figure 6.40: Proton NMR Study Comparing Bound 6.14 and Post Acidification with Dueterated TFA 344
Figure 6.41: Proton NMR of 6.14 Post Acidification with Dueterated TFA 345
Figure 6.42: : LRMS of Assembly Formed Between Compounds 6.1, 6.2, 6.11 346
Figure 6.43: LRMS of Assembly formed between compounds 6.1, 6.2, 6.13 346
Figure 6.44: LRMS of assembly formed between compounds 6.1, 6.2, 6.14 347

Chapter 1: Peptide and Unnatural Oligomer Chemistries[1]

1.1 INTRODUCTION

In the late 19th century, chemists began to explore ways to make polypeptides. Work started with simple synthesis of dimers[1, 2] Since then, peptide chemistries have been the focus of heavy optimization. Increasing the number of residues into a sequence was the goal by improving the efficiency of amide coupling.[3] In large part, this research focused on solid-phase approaches, but solution-phase chemistries remained popular.[3-6] Such work allowed the synthesis of natural peptides for biological and therapeutic applications. The more chemists delved into making natural peptides, groups began the integration of unnatural residues. Studies on these non-canonical side chains probed into their bioactivities and biomedical applications. [7] Many found use in research settings to gain better insight into many biological processes and pharmacological drug design.[8]

This kind of research also provided insight into rationally designed, controlled growth of polypeptides. Despite the research done in this field, peptide chemistry continues to be an exciting area of research. As a fundamental backbone in nature, the polypeptide will still be one of the preeminent natural scaffolds employed for designing a set of diverse compounds by changing side chain functionalities and sequence. In addition, this knowledge in linear synthetic design of peptides has inspired chemists to make unnatural oligomers.[9, 10]

The past two decades has provided a host of unnatural backbones made to study their folding properties.[11] Folding of these structures in complex arrangements was given the term "foldamer formation," and are geared to have biomimetic properties with the intended purposes of improving bioactivities of natural analogues prone to degradation *in*

[1] Adapted from T. Kimmerlin and D. Seebach, *J. Pep. Res.*, 2005, **65**, 229-260.

vivo.[7, 12] The current state-of–the-art explores foldamers in their design of building blocks, elongation of them into a sequence, characterization of higher order structures, followed by identifying any applications they could have in biological, medical, materials, and organic chemistry settings.[13]

The work presented in this dissertation is an exploration of synthetic peptide chemistry in its manipulation post-solid phase synthesis. Manipulation of peptides inspired work in appending functionalities that could induce folding of peptides with unnatural moieties. The chemistries employed for these kinds of modifications included well-established techniques known in mass proteomics and the incorporation of residues derivatized with simple modifications, such as alkylations. The overarching aim has been to find unnatural moieties that can be used for various applications like binding to target analytes.

Finally, from this research, a different way to conceptualize unnatural oligomer design is proposed. It is an approach that tries to resolve issues that include scaling-up of building blocks, improving sequence reproducibility, and arriving to a set of diverse sequences by a straightforward functionalization of the backbone. Designing this new oligomer included using dynamic, covalent bonds with well-established supramolecular interactions, and reflected the need to integrate many sub-disciplines in chemistry that include but are not exclusive to physical organic, polymer, biological, analytical, and materials chemistry. As with many projects, this work found inspiration outside of chemistry, particularly in systems biology and molecular biology. Collaborations with bioinformatcians and molecular biologists brought to light the need for supramolecular chemists to dive deeper into larger-scale production and characterization of a diverse set of molecules interacting in a confined space. To be more aware of designing chemistries that can function orthogonally with others and inducing folding of peptides and unnatural

oligomers will allow new and exciting advancements in the field. The aim should be to emulate the diversity of compounds and functions as observed in nature.

1.2 Abridged Literature Review

1.2.1 Solution-Phase Peptide Synthesis: Short Peptides Initially Made

Starting in the 1880s, chemists, like Theodore Curtius, began devising chemistries to form polypeptides. One of the first made was a sequence consisting of two glycine residues. Although, the N-terminus remained protected, this peptide preceded the more notable work of Emil Fisher who made the same sequence by hydrolyzing diketopiperazine. Fisher's work marked the beginning of peptide chemistry, starting in the early twentieth century. Both men took different approaches to making polypeptides.

Curtius relied on an azide coupling methodology, whereby he would build a sequence from the N-terminus. The next building block displaced an azide alpha to the carbonyl group. Introduction of hydrazine, followed by nitrous acid generated the azide, thus propagating the sequence (Figure 1.1). Fisher focused on acylchlorides. In the presence of free amino acids and PCl_5 in a solution of acetyl chloride, peptides could be generated. The limitations of both methodologies were the inability to isolate enantiomerically pure compounds and the difficulty to remove the protecting groups used.

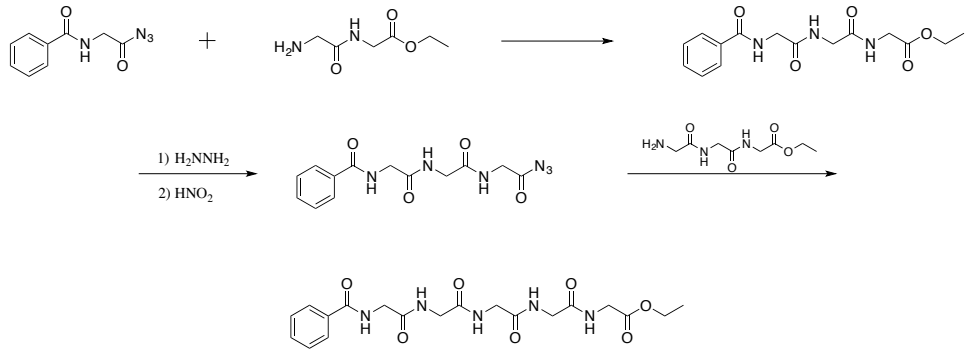
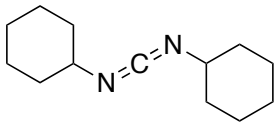


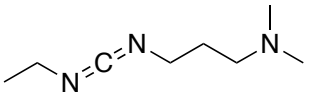
Figure 1.1: Coupling Methodology Used for Synthesis of Benzoylpentaglycine.

The following decades led to advances in the use of different protecting groups that facilitated synthesis of L-peptides. This work pioneered by chemists such as Bergman and Zervas employed carbobenzoxy (Cbz) for protection of the N-terminus. Deprotection done by hydrogen with palladium, sodium with aqueous ammonia, or hydrobromic acid in acetic acid afforded release of the amino group. With a viable means to protect and deprotect the α-amine, several naturally occurring peptides were synthesized. By the late 1950s, Carpino, McKay, and Albertson reported the boc-protecting group. This protecting group was stable to hydrogenation and basic conditions. Acidic treatment removed this protecting group and so it was orthogonal to other conditions required for other groups like Cbz[1]

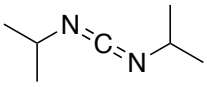
Further advances in protecting group chemistries facilitated the design and synthesis of longer sequences, such as β-Corticotropin. Along with advancements in protecting groups, amide coupling also played an important role. To minimize racemization, coupling reagents such as N,N'-dicyclohexylcarbodiimide, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, N,N'-diisoprpylcarbodiimide were employed (Figure 1.2).



N,N'-Dicyclohexylcarbodiimide



1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide



N,N'-Diisopropylcarbodiimide

Figure 1.2: Amide Coupling Reactants.

In addition to formulating protecting methodologies for the N-terminus and C-terminus, protecting groups for side chains have also been optimized extensively. Since

certain amino acid residues can interfere with the coupling process, orthogonal protecting groups that survive amide-coupling conditions have been devised. Decades of research have afforded coupling reagents and protecting groups geared to improve efficiencies in coupling while minimizing racemization.

Despite advances in solid-phase chemistries, solution-phase peptide chemistries continue to be reported. Such work strives to minimize the use of harsh conditions to generate polypeptides. Other groups have employed solution-phase strategies to minimize the costly and inefficient consumption of starting materials used in solid-phase approaches. Recent work by the Li group demonstrated chiral and achiral auxiliaries for purification of peptides in what was referred to as group-assisted purification (GAP). The authors claim these auxiliaries facilitated peptide synthesis by isolating tyrosine modified with N,N'-Di-phenyl-1,2-ethyldiamino *O*-phosphonyl groups.[14]

Many of the peptides synthesized via solution-phase approaches proposed is to have methodologies that allow for facile scale-up of peptides for research and therapeutic applications. Solid-phase approaches do not readily lend themselves for scaling up of peptides. Finding solution-phase approaches will remain an important area of research, particularly with an emphasis of devising more environmentally responsible means of synthesizing peptides.

1.2.2 Solid-Phase Peptide Synthesis: Preferred Method for Short Polyamides

Instead of relying on solution-phase approaches, which required onerous purification steps before incorporating the next amino acid, Merrifield conceptualized a new way to make peptides.[15] His approach immobilized the first residue on a solid support followed by introduction of the next amino acid building block. The linker connecting the peptide to the polystyrene support was acid labile and orthogonal to the basic conditions

for amide coupling. Mixtures of solvents and reactants along with the use of insoluble polymer made automating peptide synthesis possible.

Cleavage of peptide from the solid support coincided with deprotection of acid-labile groups protecting the side chain residues. To date, several companies offer automated peptide synthesis instruments, which rely on the chemistries developed for over a century. Recent advances have included the use of microwave reactors to improve efficiency of coupling and deprotection of the fluorenylmethyloxycarbonyl (Fmoc) group, decreasing the time to synthesize a peptide. To date, with these highly optimized systems, 10 amino acids can be added into a sequence in about an hour.

Solid supports available are numerous[16], but the most used is the preloaded Wang resin, because of its reliable cleavage and high yields. Choice of resin is important for minimizing deletion products and the accommodation of sequences longer than 20 residues.[17] Other resins are available to release the C-terminal end with different functionalities. For example, cleavage of peptide from the rink amide, releases the C-terminus as an amide. To release a thiol-terminated C-terminus, cysteamine 4-methoxytrityl resins can be used. These resins tend to be hydrophobic and the use of polar solvents or aqueous mixtures shrinks the resin, or causes them to aggregate. When aqueous mixtures are required, hydrophilic resins are available. These are known as Tentagel resins, and they are functionalized with PEG chains, improving swelling of solid supports in more polar solvents.

To date, several groups continue optimizing solid-phase chemistries, to resolve issues that include residue protecting group instability, deletion products, and overall efficiencies. Efforts have brought about a host of protecting groups tailored for solid supports. Amide coupling reagents developed for solution-phase approaches have been transferred for solid-phase chemistries. In particular, phosphonium coupling reagent, like

BOP, AOP, PyBOP, PyAOP, HBTU, HATU, HBPyU, and HAPyU, have gained popularity for automated synthesis (Figure 1.3).[18]

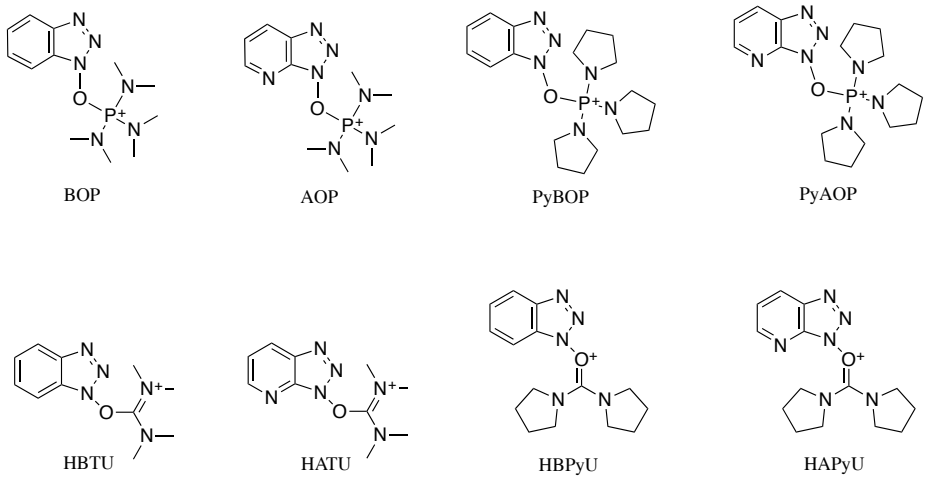
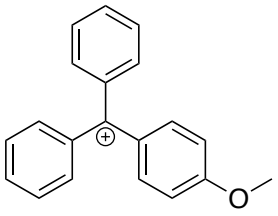


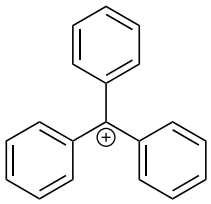
Figure 1.3: Amide Coupling Reactants Used in Solid-Phase Peptide Synthesis (SPPS)

Chemistries cleaving peptides from the solid support have also been optimized. The byproducts formed by removing acid-labile protecting groups from residues lead to reactive species, such as 2-methylpropan-2-ylium, triphenylmethylium, or (4-methoxyphenyl)dipheylmethlium (Figure 1.4). To quench these reactive byproducts, introduction of weak reducing reagents, such as triisopropylsilane (TIS), and nucleophiles, such as water, are commonly employed. These become part of the acidic cocktail to minimize side reactions that may occur with nucleophilic residues like Cysteine and Lysine.

2-methylpropan-2-ylium



(4-methoxyphenyl)diphenylmethylium



triphenylmethylium

Figure 1.4: Cationic Species Formed from Deprotecting Side Chains.

In addition to TIS or water, ethane dithiol (EDT) can be added to outcompete thiol residues from reacting with electrophilic species or from forming disulfide bonds. Other protecting groups like 4-methoxy-2,3,6-trimethylbenzensulphonyl (Mtr) for Arginine residues, different additives such as thioanisole and bromotrimethylsilane (TMSBr) are preferred. Photocleavable and oxidative cleavable resins are also available, and are used to overcome issues associated with deprotection of immobilized peptides with acid labile protecting groups and preventing cleavage from the solid support. For example, derivatives of 1,2-dimethoxy-4-nitrobenzene have been utilized as photocleavable linkers. Copper acetate can be used for oxidative cleavage of hydrazino benzoyl linkers. All these resins are aimed to improve isolation of desired peptide, while minimizing side reactions between unprotected residues and post-deprotection byproducts.

Post-synthesis these peptides are purified by High-Performance Liquid Chromatography (HPLC). This purification method remains the most commonly used way to isolate peptides. Impurities removed come from side products formed during deprotection and deletion products for longer sequences. After purification,

characterization takes place with mass spectrophotometry, and sequence information can be confirmed via fragmentation along the backbone. Because peptide chemistry is highly optimized, sequence fidelity for synthetic peptides are regarded to be reliable, especially with current instrumentation. Therefore, synthetic peptides are normally characterized by identifying exact mass to confirm the correct sequence was made. More traditional organic analytical methods like NMR are not a common technique for routine analysis of synthetic compounds. However, it is possible to obtain sequence information, but two-dimensional techniques are required.

1.2.3 Natural Residue Synthetic Peptides: Chemistry Approaches to Making Peptides

Over 7000 naturally occurring peptides are known, and many of these have been synthesized using chemistries described in previous sections.[19] These peptides play important roles as hormones, growth factors, ion channel ligands, neurotransmitters, and anti-microbial agents.[20, 21] Some of the first to be made were polypeptides such as glutathione by Harington in 1935. Vigneaud synthesized oxytocin, a naturally occurring cyclic peptide, in 1953. By the 1960s, peptides like β–corticotropin, a 39 amino acid long polypeptide, were synthesized.

Peptides with naturally occurring post-translation modifications (PTMs) have been obtained with solid-phase methods. For example, palmitoylation, myristoylation, farnesylation, and geranylation of peptides are used for anchoring peptides into the cell membrane. These peptides can be used for signaling vesicular transport, differentiation, and cell growth. Appending oligosaccharides, glycosylation, is also possible for making synthetic peptides that participate in promoting cell-cell interactions, targeting of cell, and extending protein half-life. Signal transduction can also be promoted by peptides that have been phosphorylated, and strategies to phosphorylate are reported. Biosynthetic means are

available for making these PTMs peptides, however, usually small quantities are isolated at great expense, and mixtures of sequences are obtained.[20] These mixtures make it difficult to determine which sequences have a role in biological processes. Drug candidates such as antitumor vaccines MUC1 and TLR2 glycolipopepetides have been designed with solid-phase approaches. Fragment condensation approaches in solution have also been made to make larger sequences of naturally occurring peptides. Native chemical ligation (NCL) in solution and on a solid support has been used. These chemistries rely on thioesters becoming modified with amino group attacking the carbonyl group. The amine is the α-amino group from the peptide coupled to the C-terminus. In solution, a cysteine residue is exposed and becomes part of the sequence of the large peptide formed (Figure 1.5).

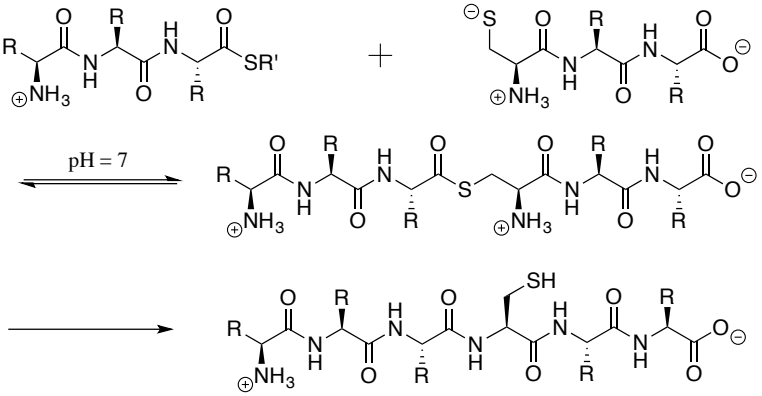


Figure 1.5: Native Chemical Ligation (NCL).

1.2.4 Unnatural Synthetic Peptides: Expanding the Residue Repertoire

It can be argued the origins of unnatural residues stem from amino acid building block synthesis. With the development of methods to make natural amino acid building blocks and modification chemistries to alter natural residues, peptide chemists also began to explore the incorporation of unnatural functionalities.[22-27] The need to modify residues

and make them stable for further modification, by clever means to make natural amino acid building blocks, provided intellectual fodder for non-canonical side chains. A common approach to introduce unnatural side chains has relied on azido lysine or alkyne-based amino acids. Triazole formation via an alkyne and copper catalyst has been used for solution-phase and solid-phase modifications. Many examples demonstrate peptides modified with alkynes altered with different peptides, fluorophores, fluorinated aromatic rings, analine derivatives, and ferrocenes.[28]

Peptides with unnatural residues have found commercial use and are lauded for their improved proteolytic stability while still retaining the relevant biological activity of their natural counterparts. Registered drugs like Abarelix, a gonadotroping releasing hormone, has an N-isoproyl Lysine. Lacosamide used to treat onset seizures and neuropathic pain is a dipeptide with a methylated serine. In conjunction with natural residues, single substitution for one amino acid can improve the efficacy of a peptide-based drug. In research settings, the Anslyn group has used modification of cysteine residues with SOX fluorophores to differentiate MAP kinases.[29]

1.2.5 Foldamers: Expanding the Backbone Repertoire for Biomimicry[9, 30]

Elongation of amino acid building blocks with methodologies developed for converting between protecting groups paved ways for chemists to consider new backbone designs that could fold in arrangements like biopolymers. Forces governing the folding of polymers include various non-covalent interactions, like hydrogen bonding, ion pairing, van der Waals and hydrophobic interactions.[31] For proteins, there is a balance between maximizing favorable side chain interactions and reducing the contacts of hydrophobic residues with the solvent. Folding, in other words, has been defined as a balance between the entropic loss, gained by restricting conformational degrees of freedom, and enthalphic

gains made by maximizing favorable interactions with solvophilic residues.[31] Keeping in mind these kinds of approaches can also be applied in the design of foldamers.

The Gellman group substituted residues with β amino acids to make α,β peptides that were more resistant to degradation compared to their natural counterparts.[32] Other unnatural building blocks employed for foldamers include peptoids, oligoureas, and aza peptides.[33] Moieties like aryl amides and phenylene ethynylene foldamers serve as examples that differ from the flexible building blocks mentioned. Higher order structure such as β-sheets and α-helices have been reported with these building blocks. With analine derivatives, double helices were possible. These unnatural oligomers have also been used for devising molecular structures that bind a target analyte by finding the most thermodynamically favorable state when bound. For example, *m*-phenylene ethylene helical hosts binding chiral guests, such as α-pinene, have been reported.

1.3 Summary

Since the late 19th century, optimizing synthetic peptide chemistry has been the focus of much research. Solution-phase approaches were first developed, but suffered from difficult purifications and racemization. The introduction of protecting groups, like Cbz and Boc, helped to minimize these issues and improve efficiencies of peptide bond formation. These advancements led to the first syntheses of naturally occurring peptides. To overcome further inefficiencies and minimize the need for purification steps, solid-phase approaches were developed. Both natural and unnatural peptides have been made with these methodologies, and synthesized peptides have found use in research and therapeutic settings. From these chemistries, a rich environment of building block design, protecting group strategies, and side chain functionalities served as inspiration for unnatural oligomer design. Like peptides, these oligomers are studied for their folding

properties in hopes of finding unnatural analogues that exhibit activities and properties found in nature.

1.4 Dissertation Outlook

The work presented in this dissertation describes utilizing well-established chemistries to forge new paths for modifying side chains, advancing proteomic technologies, and inducing folding of unnatural peptides. Extensive work done in designing and synthesizing peptides inspired new ways of conceptualizing the making of unnatural oligomers. The aim is to take the best of peptide chemistries and mitigate some of its drawbacks when considering making biomimetic analogues. Chemists, however, seeking to make foldamers should appreciate the work that began over a hundred years ago to make one of the foremost "natural foldamers." Therefore, this work focuses strongly on peptide design, and only begins to explore unnatural oligomers towards the end.

1.4.1 Chapter 2: Natural Side Chain Modifications

In this chapter, reactions known in mass proteomics were used in a sequential, linear implementation for modifying the most reactive natural residues. The modifications targeted Cysteine, Lysine, Aspartate, Glutamate, the N-, and C-termini. Modification of side chains first occurred in solution and was transitioned to the solid phase for labeling model peptides KDYWEC and KDYWE. After learning these modifications, work on incorporating functional handles for selective labeling of residues with fluorophores became a focus.

1.4.2 Chapter 3: Fluorescent Peptides Used for a Single-Molecule Peptide Sequencing Platform

In collaboration with the Marcotte group, unnatural peptides modified with fluorescent probes were made. These peptides were used in proof-of-concept studies for a new peptide-sequencing platform aimed to extract sequence information at the single

molecule level. The chemistries required for making peptides with appropriate controls required expanding our understanding of synthetic peptide design. For example, synthesis of fluorescently labeled building blocks and incorporation into automated synthesis provided lessons of introducing unnatural functionality during solid-phase synthesis. Or, labeling of peptides at the N-termini with dyes on the solid support, demonstrated the straightforward methodologies to incorporate, complex unnatural functionality post solid phase synthesis.

1.4.3 Chapter 4: Secondary Amine Residue Peptides Alkylated with Boronic Acids

Incorporation of unnatural side chain functionality of peptides became an important goal in our synthetic strategy for introducing non-canonical side chain functionalities. To do so, we developed three approaches to incorporate secondary amines into our sequences. These methodologies included using solution-phase and solid-phase chemistries followed by alkylation of these amines to form boronic acid peptides.

1.4.4 Chapter 5: Secondary Amine Residue Peptides Alkylated with Bipyridines to Induce Folding

Secondary amines were used again for appending bipyridines off the amide backbone. These unnatural moieties in the presence of Fe^{2+} and free bipyridines in solution caused cyclization of these synthetic peptides. After forming higher order structures, these cyclic peptides were introduced into a solution of hydrazone forming peptides and studied for higher order structures formed.

1.4.5 Chapter 6: Modular Oligomer Design Achieved by Using a Guanidinium, Boronic Acid, and Dione

A new oligomer was designed from literature reported side chain modification strategies. Diones and boronic acids form a tri-component assembly with guanidyl groups like Arginine, and this reaction was used for the synthesis of a synthetic oligomer

incorporating terpyridines complexing about Zn^{2+}. The design of this oligomer arose from the principles of "modular oligomer design," a term conceived from research done in peptide chemistry. The term aims to emphasize the need for designing synthetically accessible building blocks, the use of simple and reliable "linking" strategies for oligomer formation, and the ability to modify the backbone post oligomer formation. With these principles, peptide and unnatural oligomer chemistry will facilitate synthesis of foldamers. In addition, the Anslyn group believes modular oligomer design will direct efforts to improve the scaling-up of building blocks, improve the atom economy of oligomer formation, and decrease time spent before characterizing these structures. The hope will be to increase the scale, reproducibility, and structure complexity of synthetic structures designed to emulate macromolecules in nature.

1.5 References

1. T. Kimmerlin and D. Seebach, *J. Pep. Res.*, 2005, **65**, 229-260.
2. G. Jung and A. G. Beck-Sickinger, *Angew. Chem. Int. Ed.*, 1992, **31**, 367-383.
3. Y. Tsuda and Y. Okada, in *Amino Acids, Peptides and Proteins in Organic Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2010, DOI: 10.1002/9783527631803.ch6, pp. 201-251.
4. L. A. Carpino, S. Ghassemi, D. Ionescu, M. Ismail, D. Sadat-Aalaee, G. A. Truran, E. M. E. Mansour, G. A. Siwruk, J. S. Eynon and B. Morgan, *Org. Process Res. Dev.*, 2003, **7**, 28-37.
5. C. J. Forsyth, *Nat. Chem.*, 2010, **2**, 252-254.
6. R. A. Khan, *Chem. Biol. Drug. Des.*, 2016, **88**, 884-888.
7. G. Luca, M. Rossella De and C. Lucia, *Curr. Pharm. Design*, 2010, **16**, 3185-3203.
8. D. B. F. Johnson, J. K. Takimoto, J. Xu, L. Wang and T. P. Begley, in *Wiley Encyclopedia of Chemical Biology*, John Wiley & Sons, Inc., 2007, DOI: 10.1002/9780470048672.wecb585.
9. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore, *Chem. Rev.*, 2001, **101**, 3893-4012.
10. C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat. Chem. Biol.*, 2007, **3**, 252-262.
11. T. A. Martinek and F. Fulop, *Chem. Sov. Rev.*, 2012, **41**, 687-702.
12. M. T. Weinstock, J. N. Francis, J. S. Redman and M. S. Kay, *Pep. Sci.*, 2012, **98**, 431-442.
13. in *Foldamers*, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, DOI: 10.1002/9783527611478.fmatter, pp. I-XXII.

14. J. Wu, G. An, S. Lin, J. Xie, W. Zhou, H. Sun, Y. Pan and G. Li, *Chem. Comm.*, 2014, **50**, 1259-1261.
15. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149-2154.
16. P. H. H. Hermkens, H. C. J. Ottenheijm and D. Rees, *Tetrahedron*, 1996, **52**, 4527-4554.
17. F. Albericio and J. Tulla-Puche, in *The Power of Functional Resins in Organic Synthesis*, Wiley-VCH Verlag GmbH & Co. KGaA, 2009, DOI: 10.1002/9783527626175.ch1, pp. 1-14.
18. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606-631.
19. K. Fosgerau and T. Hoffmann, *Drug Discov. Today*, 2015, **20**, 122-128.
20. J. M. Palomo, *RSC Adv.*, 2014, **4**, 32658-32672.
21. R. Karstad, G. Isaksen, B.-O. Brandsdal, J. S. Svendsen and J. Svenson, *J. Med. Chem.*, 2010, **53**, 5558-5566.
22. A. Stevenazzi, M. Marchini, G. Sandrone, B. Vergani and M. Lattanzio, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 5349-5356.
23. A. Ahmad Fuaad, F. Azmi, M. Skwarczynski and I. Toth, *Molecules*, 2013, **18**, 13148.
24. K. E. Schwieter and J. N. Johnston, *J. Am. Chem. Soc.*, 2016, **138**, 14160-14169.
25. D. A. Dougherty and E. B. Van Arnam, *Chem. Bio. Chem.*, 2014, **15**, 1710-1720.
26. C. Han and J. Wang, *Chem. Phys. Chem.*, 2012, **13**, 1522-1534.
27. B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem.*, 1996, **4**, 5-23.
28. H. Li, R. Aneja and I. Chaiken, *Molecules*, 2013, **18**, 9797.
29. D. Zamora-Olivares, T. S. Kaoud, J. Jose, A. Ellington, K. N. Dalby and E. V. Anslyn, *Angew. Chem. Int. Ed.*, 2014, **53**, 14064-14068.
30. S. H. Gellman, *Acc. Chem. Res.*, 1998, **31**, 173-180.
31. I. Saraogi and A. D. Hamilton, *Chem. Soc. Rev.*, 2009, **38**, 1726-1743.
32. W. S. Horne, L. M. Johnson, T. J. Ketas, P. J. Klasse, M. Lu, J. P. Moore and S. H. Gellman, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 14751-14756.
33. J. S. Laursen, J. Engel-Andreasen and C. A. Olsen, *Accts. Chem. Res.*, 2015, **48**, 2696-2704.

Chapter 2: Solution-Phase and Solid-Phase Sequential, Selective Modification of Side Chains in KDYWEC and KDYWE as Models for Usage in Single-Molecule Protein Sequencing[2]

2.1 INTRODUCTION

2.1.1 Need for Selective Labeling of Amino Acid Residues for Emerging Single-Molecule Sequencing Technologies

Recent theoretical work on single-molecule peptide or protein sequencing suggests that modifying proteins with amino acid side chain specific labels, and determining the order of the subset of modified amino acids (e.g. via single molecule fluorescence microscopy in combination with Edman sequencing[1] or nanopore-based sequencing)[2] can be sufficient to identify proteins in complex mixtures at the single molecule level. Such a technology offers the potential for multiple orders of magnitude improvements in sensitivity and throughput over existing approaches, but still faces barriers for its practical implementation. A key barrier is the development of a side-chain specific labeling scheme capable of modifying multiple amino acid types with specific and detectable labels, such as residue specific fluorophores. In such a scheme, only the positions of the fluorescent amino acids are determined for a peptide or protein, which is then identified by comparison with the sequences expected based upon that organism's known genome sequence. Thus, more amino acid-specific modifications translate directly into richer sequence information about a given peptide or protein, ultimately allowing for greater coverage of that organism's proteome. Computational modeling suggests that schemes incorporating

[2] **Hernandez, Erik**; Swaminathan, Jagannath; Marcotte, Edward M.; Anslyn. Eric V. "Peptide Modification Chemistry: Solution-phase and Solid-phase Sequential, Selective Modification of Side Chains in KDYWEC and KDYWE." *New Journal of Chemistry*. **2017**, *41*, 462-469. Designed peptides and did labeling chemistries proving sequential, selective modification.

modifications to cysteine, lysine, tryptophan, aspartic, and glutamic acid residues can in principle provide a high coverage of the human proteome.[1]

2.1.2 Chemistries Available for Selective Modification of Peptides

Fortunately, work developed for mass spectrometry proteomics has provided optimized chemistries that selectively target residues for modification. These techniques have gained widespread use and are employed to understand biological processes such as expression, post-translation modifications, and protein interactions. These labeling chemistries are highly regarded for their efficiency and low cross-reactivity. Side chain specific protocols are routinely used to tag and modify proteins with high selectivity, where little to no cross-reactivity has previously been established.[3] For example, kits are commercially available for targeting cysteine and lysine. The cysteine is modified with an iodoacetamide, followed in the same-pot by selective modification of lysine using *o*-methylisourea hemisulfate.[4] Acylation and reductive alkylation are also employed to modify both Ne-amines and N-termini.[5] Cross-modification of threonine, serine, and tyrosine can occur with acylation and alkylating conditions.[6] Recently, peptide amines have been modified via reductive methylation preventing cross-reactivity with alcohol and phenol residues. Once these amines were modified, the Smith group achieved global modification of aspartate and glutamate via amidation with amine-containing compounds. Furthermore, studies with less abundant side chains have been explored. For example, Horton, Koshland, and Scoffone demonstrated the modification of tryptophan under acidic conditions using 2-hydroxy-5-nitrobenzyl bromide and dinitrophenylsulfenyl chloride.[7–9] More recently, studies of the N-terminus using pyridinecarboxyaldehydes have expanded the chemical repertoire available for selective targeting of functional groups.[10]

2.1.3 Scientific Aim: Modify Most Reactive Amino Acids on Model Peptides

However, missing in these studies is a route integrating all these selective modifications into a sequential protocol. Such a route, taking advantage of these techniques, could have potential applications not only in these emerging single-molecule technologies, but also in protein/peptide mass spectrometry studies. In addition, devising a generalized sequential route consisting of the selective modification of side chains also has applications for synthetic peptide and protein design.[11–13]

Herein, we describe studies that put this sequential strategy described in scheme 2.1 and 2.2 into practice. As described, a series of selective modification steps for KDYWEC (2.1) and KDYWE (2.2) was achieved. Modification studies were initially

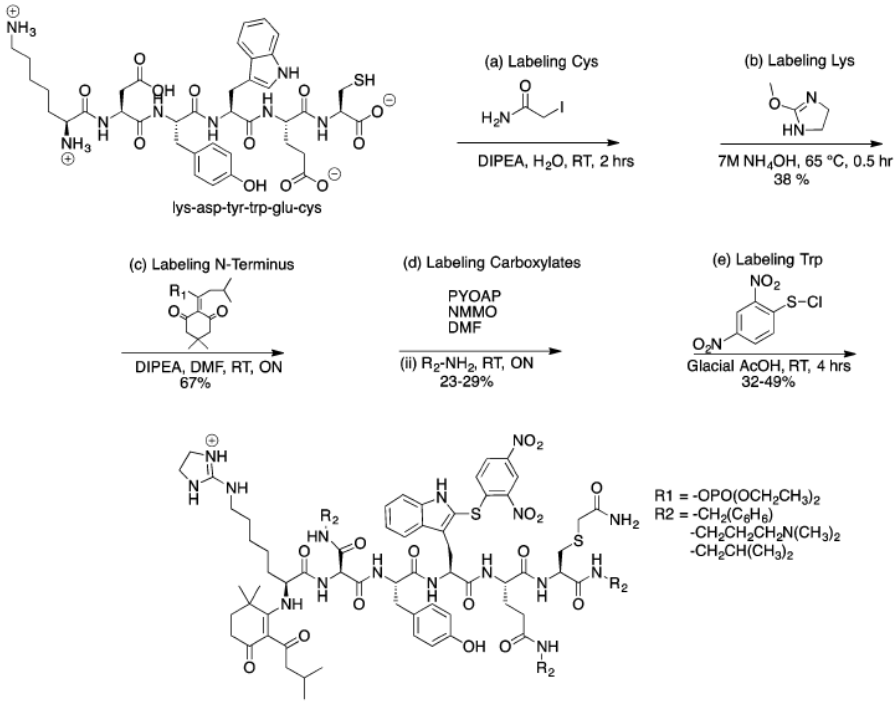


Figure 2.1: Selective modification route in solution.

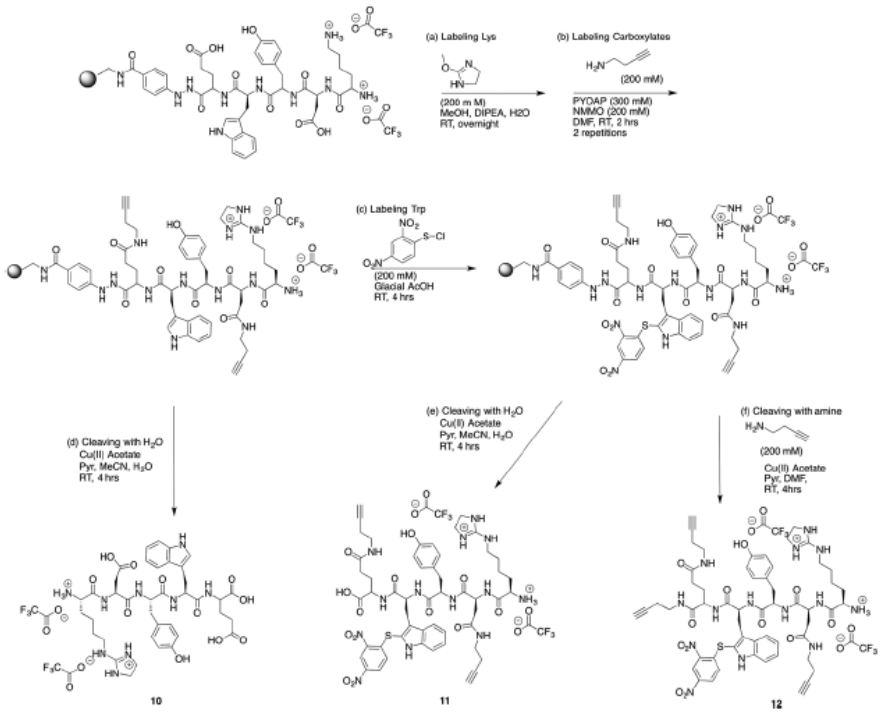


Figure 2.2: Solid-phase modification of KDYWE. (a) Modification of lysine was done similarly for immobilized KDYWEC. (b) Two repetitions were performed to drive reaction to completion. (c) Modification of tryptophan. (d and e) Cleavage with water releases C-terminus as an acid. (f) Cleavage with amine functionalized C-terminus with an alkyne.

performed in solution-phase targeting cysteine, lysine, the N-terminus, aspartic, glutamic acid, and tryptophan. For ease of identification and purification, solution-phase studies were performed. These chemistries were transitioned to the solid-phase. Both peptides proved to be suitable models for the ultimate use in single-molecule peptide/protein sequencing.

It was presumed that global modification of several amino acids of different classes should be possible if the appropriate conditions and the right sequence of derivatization steps were used. For example, by using iodoacetamide, guanidination reagents, amide-coupling reagents, and aromatic-targeting electrophilic reagents, we may selectively functionalize cysteine, lysine, aspartic and glutamic acids, and tryptophan, respectively, if used in the proper protocol. Minimizing cross-reactivity between each step was to be achieved if the nucleophilicity and pK_a of each side chain, as well as the reactivity of modification of reagents, reactions times, and temperature are all considered. Good nucleophiles, such as the thiol in cysteine or the amine in lysine, and the N-terminus, should be targeted first. Selective modification of cysteine between pH 7–8 is possible, while modification of amines is possible at a higher pH.[14] Because guanidinating reagents are selective for amines, distinguishing between the Ne-amine of lysine and the terminal a-amine was possible. As described herein, a different modification of reagent is required for each kind of amine. Once the most nucleophilic sites were modified, we tested if the carboxylate side chains could then be targeted followed finally by modification of tryptophan.

Characterization of the resulting peptides in nearly all these examples is performed only with HPLC for purity and mass spec for identity due to the minimal quantity of material produced. As is the case with the studies reported herein, the quantity of material is not amenable to ^{1}H and/or ^{13}C NMR spectroscopy. Our goals herein were to create a protocol for sequential modification of numerous side chains using known selective derivatizations, and to verify that each gave the expected products after step-by-step linear implementation.

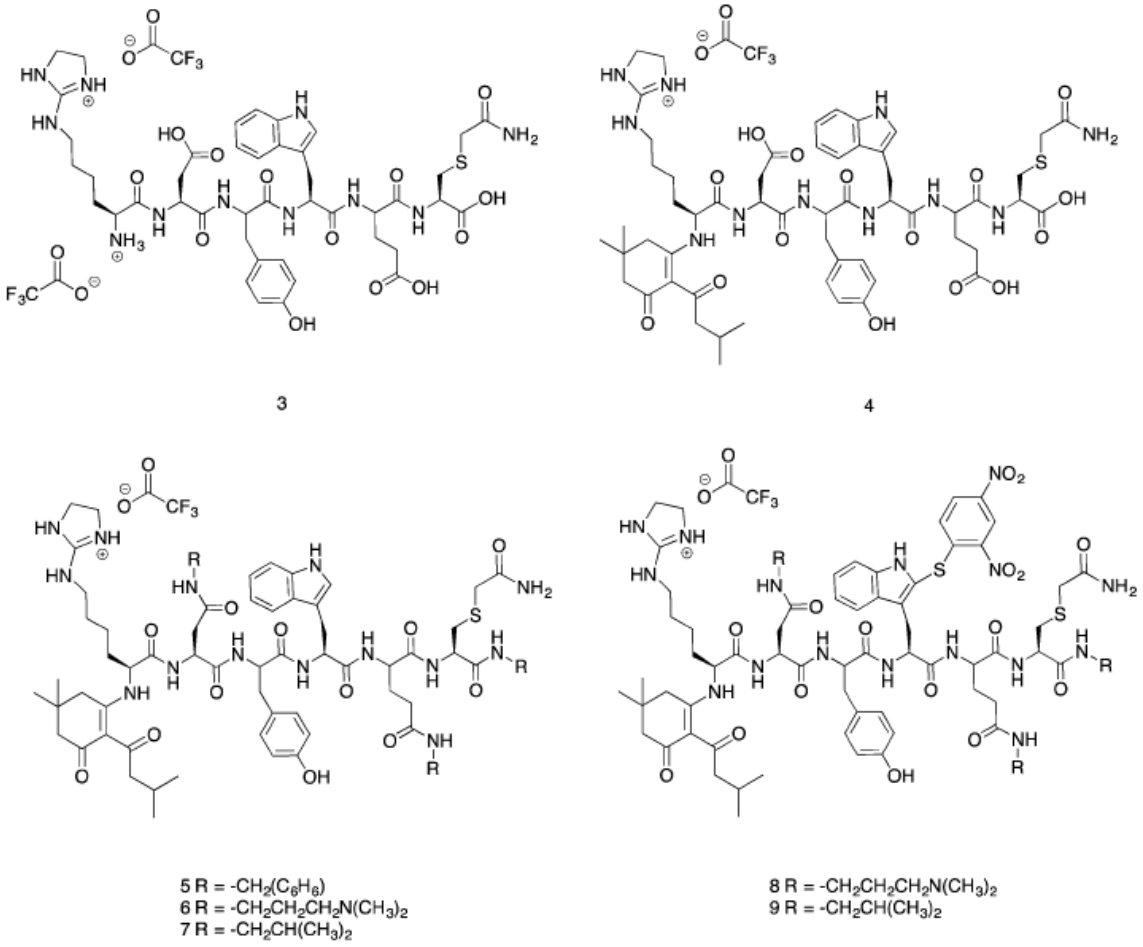


Figure 2.3: Modified model peptide intermediates.

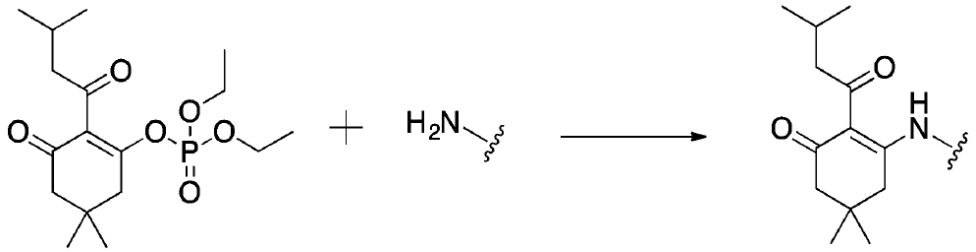
2.2.1 Solution-Phase Sequential, Selective Modification of KDYWEC

In the experimental design we implement here, we are exploiting reagents that have already been proven to have high selectivity for their individual amino acid side chains. Our goal was to direct the derivatizations by modifying experimental conditions and then use each derivatization step sequentially, which, to our knowledge, has not been previously accomplished. The order of steps in scheme 2.1 took into consideration the nucleophilicity and acid/base-dependent reactivity of the target side chains in KDYWEC. This peptide was synthesized to contain the most reactive natural amino acids. The sulfhydryl group in

cysteine is the most nucleophilic, and is prone to oxidation and disulfide bond formation. To ensure selectivity in future modification steps, cysteine
was first alkylated with iodoacetamide, forming a stable thioether. Maintaining a pH between 7–8 ensured the amines remained protonated, thus limiting the possibility of undesired alkylation. Subsequently, the pH was raised to 11 and 2-methylthio-2-imidazoline hydroiodide (MDI) was introduced. Modification of the Ne-amine occurred in 24 minutes when heated to 50 1 °C. Longer reaction times increased the extent of N-terminal modification. The Cys and Lys modification steps were performed in one-pot. The yield of peptide 3 after purification was 38%. Because the lysine was modified while heating under basic conditions, the thioether and guanidinium group were considered to be stable in future derivatization steps.

Of the remaining nucleophilic sites, the N-terminus was first targeted. Protection of the N-terminus was required prior to modification of aspartate, glutamate, and the C-terminus. If not, concatenation of peptides could occur during amidation. The modification conditions of the N-terminus also required a group compatible to both basic and acidic conditions in subsequent derivatization steps. Literature accounts have reported using 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl to protect amines during peptide synthesis. This protecting group is stable to highly basic and acidic conditions, and is removed under hydrazinolysis conditions.[15] However, refluxing overnight to efficiently add the protecting group is common. Heating overnight was undesired so as to minimize observed degradation. Thus, we devised the use of 5,5-dimethyl-2-(3-methylbutanoyl)-3-oxocyclohex-1-en-1-yl diethyl phosphate (phos-DOD) as an alternative. Diethyl phosphate was anticipated to be a better leaving group, thereby facilitating the reaction. This compound was formed with chloro diethyl phosphate *in situ*, followed by incubation with

a basic solution of peptide 3 overnight. Post-purification the yield of peptide 4 was 67%. Protection of the N-terminus thus occurred as shown in eqn. 2.1.



eq. 2.1

Once the most nucleophilic sites in the model peptide were modified, the carboxyl groups were targeted. Amidation has been used for derivatization of aspartate, glutamate, and the C-terminus.[16] Unlike the modification of lysine and the N-terminus, distinguishing among these target side chains was not possible. Also, because there were three sites for reaction, an efficient modification approach was necessary. Highly efficient, global modification using (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) and N-methylmorpholine (NMM) has been reported.[6] Peptide 5, amidated with benzylamine dissolved in MeCN/H_2O mixtures, and purification of the resulting modified peptides was possible. Yield for peptide 5 was 29%. To improve the solubility of the model peptide, 3-dimethylaminopropylamine (DMAPA) was used to increase the overall positive charge under acidic conditions. Purity checks of the peptide demonstrated an approximate 10-minute difference in elution when comparing peptide 5 and 6. For peptide 7, isobutylamine was explored as an alternative, believed to form a peptide less hydrophobic than peptide 5, but more hydrophilic than peptide 6. Peptide 7 readily dissolved in MeCN/H_2O mixtures, but coeluted with an impurity characterized by LCMS as m/z 313.4. The impurity was removed after synthesizing peptide 8.

Tryptophan was the remaining target. As a less abundant amino acid in nature, the ability to identify this side chain can be informative for determining the protein origin of peptides in proteomic studies.[17] In synthetic peptide design, incorporating an additional site for derivatization increases the repertoire of side chains to modify. Therefore, devising a selective modification strategy incorporating modification of tryptophan was seen as important. Cysteine reacting with sulfenyl chlorides has been reported.[18] As with the other modification steps, modification of cysteine prior to that of tryptophan was chosen. However, under acidic conditions the tryptophan could be selectively targeted in the presence of unprotected N-terminus and lysine. The advantage to modification of the tryptophan last was the relative ease of the reaction. Peptides 5 and 6 readily dissolved in glacial acetic acid, and the reaction occurred in 4 h at RT. 2,4-Dinitrobenzenesulfenyl chloride (DBSC) was a chromophore and the peptides could also be monitored at 330 nm. Yields were 32% and 49% for peptides 8 and 9, respectively.

2.2.2 Solid-Phase Sequential, Selective Modification of KDYWE

The ultimate application in single molecule sequencing of our derivatization protocol will employ immobilization of the peptides on solid supports, and subsequent attachment of the peptides to surfaces via the thiols on cysteine. Thus, in a protocol that employs solid supports for derivatization, cysteine residues do not need to be derivative, because they will be the points of attachment to the surface. Hence, we removed the cysteine of the peptide we studied in solution, resulting in KDYWE, for our solid-phase synthesis studies.

Efforts to modify the amino acids on solid phase supports were explored after the sequence of targeting side chains had been successfully demonstrated in solution. Synthetic peptides have been commonly modified when immobilized on a solid support, usually at reactive side chains such as lysine.[19] Requirements for successful solid-phase reactions

include making sure each step is highly selective. Further, the reagents must be able to diffuse into the resin to reach sites for reaction. A high concentration of starting material in the bulk solution ensures a concentration gradient is formed for reactants to diffuse.[20] Inherent in this study was devising an approach that selectively modified target side chains in a sequential fashion. Therefore, high specificity was required. Literature, and the work presented here, has demonstrated that excess reagent can be used while maintaining selectivity. The final requirement for solid-phase studies was using a resin that would not cleave with acid or base. 4-Fmoc-hydrazinobenzoyl resin AM was selected, because literature accounts describe the stability towards strong acids and bases. Peptides immobilized on this resin were isolated after oxidative cleavage with Cu(II) and base.[21,22] Fig. 3 summarizes the modification of reactions performed on the solid support for peptide KDYWE. The first side chain targeted was the lysine. Two changes were made from the solution approach. The reaction time was longer. The immobilized peptide was incubated overnight with MDI. A solution of MeOH/DIPEA/H_2O (7 : 2 : 1) (v/v/v) was used instead of a solution of NH_4OH. Overnight incubation and the use of DIPEA have been reported in the literature.[23] The doubly modified peptide was not observed after an overnight reaction at RT, however, extending the reaction time to 48 h led to doubly modified peptide. Selectivity for the Ne-amine can be explained due to inductive and steric effects. The N_ε amine in lysine is part of a hydrocarbon side chain and not adjacent to an electron-withdrawing amide group, and thus, the lysine side chain amine is more nucleophilic than the a-amine. Furthermore, the N-terminal amine is closer to the amide backbone, impeding MDI due to sterics. The same inductive and steric effects played a role when modifying KDYWEC in solution phase. However, lowering the reaction temperature from 60 °C to RT made these effects more pronounced. A protection step of the N-terminus was not performed. One reason was to discover if in the presence of excess amine, the carboxylates

would be modified without concatenation to this terminal amine. A second goal was to check if the number of modification steps could be reduced, leaving the terminal-amine unmodified for future reactions. This approach could provide synthetic flexibility by diversifying the kinds of reactions performed at the N-terminus once the peptide is cleaved from the resin. The amine used in solid-phase synthesis differed from that of the solution-phase studies. 1-Amino-3-butyne had an alkyne group that could also provide sites for derivatization via Huigen–Sharpless.[24] The same coupling reactants PyAOP and NMM were employed for solid-phase studies. Two repetitions ensured all carboxylates were modified. Cleavage of the peptide was performed using a catalytic amount of Cu(II) and a mixture of MeCN/H_2O/Pyr. To a different batch of resin, the lysine and carboxylates were also modified. Tryptophan was modified in a similar fashion as in solution, four hours at RT.[25] Two different cleavage conditions were tested for the model peptide after the target side chains were modified. The first condition was water, liberating a carboxylate at the C-terminus. Peptide 11 was thereby isolated with a HPLC purified yield of 4%. Additionally, a nonaqueous condition in the presence of a nucleophile could also be employed to cleave the peptide. 1-Amino-3-butyne was the nucleophile used, liberating peptide 12 with an HPLC purified yield of 5%. The peptide could also have been cleaved with a different nucleophile diversifying the functional groups, further differentiating between the C-terminus and carboxylate side chains. Isolating peptide 12 required extra washes with DMF, because solubility in H_2O/MeCN was reduced once an alkyne was introduced at the C-terminus. Initially, the peptide was rinsed with MeCN, and LCMS data of the extract did not indicate the presence of desired product. Once rinsed with DMF and the solvent removed, peptide 12 was observed.

2.3 Conclusions

A sequential and selective scheme using common mass modifications was developed for derivatizing peptides as a model for ultimate use in single-molecule sequencing studies, helping to overcome a key barrier identified by previous computational models of single molecule protein sequencing.[1,2] Selective targeting of side chains in KDYWEC was achieved in solution-phase. Most of the side chains, as well as the N-terminus and C-terminus were modified. No heating was required to target N-terminal amine with 4,4-dimethyl-2,6-dioxocyclohex-1-ylidene when using Phos-DOD. Selective modification also occurred in solid-phase studies for KDYWE. Modification of all the target side chains was possible while omitting any reaction at the a-amine. Oxidative cleavage of the resin provided flexibility to choose between releasing modified or unmodified C-terminus. The use of 1-amino-3-butyne as the carboxylate-modification of reagent introduced further functionality that could be exploited in future reactions. Such an approach can also have applications for peptide modification studies and novel synthetic peptide design. Other tags, like fluorescent probes, can be designed to have the same functional handles presented in this paper. Work on the design, synthesis, and selective modification of peptides using these dyes will be discussed in future publications.

2.4 Experimental

2.4.1 General Materials

For automated, Fmoc amino solid-phase peptide synthesis, OtBu (Asp, Glu), Boc (Lys, Trp), tBu (Tyr) were used. Fmoc protected amino acids were purchased from Novabiochem (USA) and AAPPTec (USA). Fmoc-Cys(Trt)-Wang resin (100–200 mesh) and 4-Fmoc-hydrazinobenzoyl resins AM Novagel™ were purchased from Novabiochem (USA). Other chemicals used for automated, solid-phase peptide synthesis were purchased from Fisher Scientific and Sigma-Aldrich. Reagents used for selective modification of studies were

iodoacetamide (IA), 2-methylthio-2-imadazoline hydroiodide (MDI), sodium methoxide, diethylchlorophosphate, 2-(3-methylbutyryl)-5,5-dimethyl-1,3-cyclohexandione, benzylamine (BA), isobutylamine, 3-dimethylaminopropylamine (DMAPA), 1-amino-3-butyne (AB), (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP), N-methylmorpholine
(NMM), and 2,4-dinitrobenzenesulfenyl chloride (DBSC). All chemicals were purchased from Sigma-Aldrich.

2.4.2 General Instrumentation

A Prelude peptide synthesizer (Protein Technologies, Inc.) was used for automated-solid phase synthesis. Preparative HPLC purification of peptides was performed using an Agilent Zorbax SB-C18 Prep HT column 21.2 x 250 mm; 10 ml min^{-1}, 5–95% MeCN (0.1% TFA) in 90 min. Analytical HPLC characterization of peptides was performed using an Agilent Zorbax column 4.6 x 250 mm; 1 ml min^{-1}, 5–95% MeCN (0.1% TFA) in 40 min (RT). An Agilent Technologies 6530 Accurate Mass
QTofLC/MS was used for high-resolution mass spectra of purified peptides. Solvents used were HPLC grade.

2.4.3 General Solid-Phase Model Peptide Synthesis

KDYWEC was synthesized using Fmoc-Cys(Trt)-Wang resin (0.57 mmole g^{-1}, 100 mmole) by sequential coupling of N_{α}-Fmoc-amino acid (0.1 M, 1.5 ml) in DMF in the presence of N,N,N,N-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU, 0.15 M, 1.0 ml) and DIPEA (0.2 M, 0.5 ml) with a reaction time of 30 minutes at room temperature. A total of three repetitions were performed for each amino acid building block. DMF (3 ml, 3 min, 3x) and DCM (3 ml, 3 min, 3x) washes were done before each repetition. After incorporation of the third amino acid, a 0.8 M LiCl wash step was performed after swelling with DCM (3 ml, 3 min, 3x). Post synthesis, resin was washed

with glacial AcOH (5 ml, 3x), DCM (5 ml, 3x), and MeOH (5 ml, 3x). The resin was placed under vacuum overnight. The peptide was cleaved from the resin using trifluoroacetic acid (TFA), triisopropylsilane (TIS), 1,2-ethanedithiol (EDT), and nanopure water (94 : 1.0 : 2.5 : 2.5), and precipitated with diethyl ether at 0 °C. No HPLC purification of the crude peptide was necessary. KDYWE was synthesized using 4-Fmoc-hydrazinobenzoyl resin AM Novagelt. TFA, TIS, and nanopure water were used (95 : 2.5 : 2.5) to deprotect the side chains, and the peptide remained immobilized on the solid support.

2.4.4 Solution-Phase Modification Studies of KDYWEC

Modification of Cysteine with iodoacetamide. Peptide 1 (75 mmole) was dissolved in 0.4 ml of nanopure water. A solution consisting of 0.37 ml of MeOH/Pyr/TEA/nanopore H_2O (7/1/1/1) (v/v/v/v) was introduced (adjusting to pH 8), followed by addition of iodoacetamide (97 mmole). The reaction was incubated for 2 h at RT. Modification of lysine with 2-methoxy-4,5-dihydro-1Himidazole (3). In the same pot, 0.5 ml of a 7 N solution of NH_4OH was added, followed by introduction of MDI (750 mmole). The reaction mixture was incubated for 24 min at 65 °C, followed by introduction of TFA (0.3 ml) at 0 °C. The crude peptide was prepared for preparative HPLC using an Extract Clean C18 500 mg/4 ml solid phase extraction column (SI). The peptide was purified using preparative HPLC, and the organic solvent in the peptide fraction was removed via rotary evaporation. Aqueous remnants were frozen at -78 °C and lyophilized overnight. Purified yield: (29 mmole) 38%. High-res MS: found m/z 968.39360, calcd 968.39310 $(M + H)^+$; found m/z 966.37880, calcd 966.37850 $(M - H)^-$.

Modification of the N-terminus with 5,5-dimethyl-2-(3-methylbutanoyl)-3-oxocyclohex-1-en-1-yl diethyl phosphate (Phos-DOD) (4). Peptide 3 (12 mmole) was dissolved in 0.1 ml of nanopure water, followed by dilution with 0.2 ml of MeCN. To the solution, 0.12 ml of 7/2/1 MeOH/TEA/H_2O (v/v/v) was introduced. A solution of Phos-

DOD (SI) (18 mmole) was introduced. The solution was incubated overnight at RT. The peptide was purified using preparative HPLC. Organic solvent in peptide fraction was removed via rotary evaporator. Aqueous remnants were frozen at -78 °C and lyophilized overnight. Purified yield: (8 mmole) 67%. High-res MS: found m/z 1174.52380, calcd 1174.52380 $(M + H)^+$; found m/z 1172.50750, calcd 1172.50920 $(M - H)^-$.

Modification of the carboxylate side chains and C-terminus with benzylamine (BA) (5). Peptide 4 (51 mmole) was dissolved in 0.2 ml of 3/1 MeOH/H_2O (v/v). In a separate vial, benzylamine (1.3 mmole) was dissolved in 0.1 ml of MeCN, followed by addition of NMM (1.0 mmole). The BA/NMM solution was introduced to the peptide solution, followed by addition of solid PyAOP (0.51 mmole) and anhydrous HOBt (0.56 mmole). 0.1 ml of MeCN was introduced to improve the solubility of PyAOP/HOBt. The solution was incubated for a total of 4 h at RT. The peptide was purified using preparative HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation. Aqueous remnants were frozen at -78 °C and lyophilized overnight. Purified yield: (15 mmole) 29%. High-res MS: found m/z 1441.71230, calcd 1441.71260 $(M + H)^+$; found m/z 1439.69600, calcd 1439.69800 $(M - H)^-$.

Modification of the carboxylate side chains and C-terminus with 3-dimethylaminopropylamine (6). Peptide 4 (11 mmole) was dissolved in 0.2 ml of dry DMF. DMAPA (1.6 mmole) and NMM (1.4 mmole) were combined in a separate vial. The amine/NMM solution was introduced to the peptide solution, followed by addition of solid PyAOP (1.9 mmole). The solution was incubated for 24 h at RT. The sample was placed in a centrifugal evaporator for 21 h at 35 1 °C. The resulting oil was dissolved in 1.5 ml of 2/1 H_2O/DMF (v/v), and purified by prep HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation. Aqueous remnants were frozen at -78 °C and

lyophilized overnight. Purified yield: (2.4 mmole) 23%. High-res MS: found m/z 1426.83900, calcd 1426.83920 $(M + H)^+$.

Modification of the carboxylate side chains and C-terminus with isobutylamine (7). Isobutylamine (60 mmole) and NMM (excess) were combined in a separate vial to make 0.1 ml solution in DMF. Amine/NMM solution was introduced to peptide 4 (20 mmole), followed by introduction of solid PyAOP. The solution was incubated for 3 h at RT, followed by quenching with 1 ml of H_2O. The solution was placed in centrifugal evaporator for 14 h at 35 °C. The residual oil was dissolved in 1.5 ml of 1/1 H_2O/MeCN (v/v) and purified via prep HPLC. An impurity and the desired compound both eluted at the same time. The peptide was therefore taken directly to the subsequent modification of tryptophan without further purification.

Modification of tryptophan in peptide 6 (8). Peptide 6 (19 mmole) was dissolved in 1 ml of glacial acetic acid, followed by introduction of 2,4-dinitrobenzenesulfenyl chloride (57 mmole). The reaction was shaken for 4 h at RT. Glacial acetic acid was removed by rotary evaporation. The residual film was dissolved in 1/1 MeCN/H2O (v/v), and purified via preparative HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation. Aqueous remnants were frozen at
-78 °C and lyophilized overnight. Purified yield: (6.4 mmole) 32%. High-res MS: found m/z 812.91050, calcd 812.91000 $(M + 2H)^{2+}$; found m/z 1622.79650, calcd 1622.79810 $(M - H)^-$.

Modification of tryptophan in peptide 7 (9). Peptide 7 (6.2 mmole) was dissolved in 1 ml of glacial acetic acid, followed by introduction of 2,4-dinitrobenzenesulfenyl chloride (19 mmole). The reaction was shaken for 4 h at RT. The peptide was purified using preparative HPLC. The organic solvent in the peptide fraction was removed via rotary evaporator. Aqueous remnants were frozen at -78 1 °C and lyophilized overnight.

Purified yield: (6.4 mmole) 49%. High-res MS: found m/z 769.37050, calcd 769.37020 (M + 2H)$^{2+}$; found m/z 1535.71420, calcd 1535.71850 (M - H)$^{-}$.

2.4.5 Solid-Phase modification studies of KDYWE

Before and after each modification step, the resin was washed with DMF and DCM (3 ml, 3 min, 3x). Resins were placed under high vacuum overnight before cleavage at each step. Cu(II)Ac (0.3 mmole) was dissolved in 3 ml 45/45/10 MeCN/H_2O/ Pyr (v/v/v). The copper acetate solution was introduced to the dried resin and incubated for 4 h at RT to cleave peptide. This solution was removed from the resin and collected, followed by washing with 1/1 MeCN/H_2O (v/v) (1 ml, 3 min, 3x); washes were collected.

Modification of the lysine with 2-methoxy-4,5-dihydro-1Himidazole in 2. To the swollen resin (130 mg, 0.66 mmole g^{-1}), 3 ml of a 200 mM solution of 2-methoxy-4,5-dihydro-1H-imidazole in 7/2/1 MeOH/DIPEA/H_2O (v/v/v) was added. The resin was incubated overnight at RT.

Modification of the carboxylates and C-terminus (10). 1-Amino-3-butyne (0.61mmole) was dissolved in NMM (0.45mmole), and the mixture was diluted with 1ml of DMF. PyAOP (0.40 mmole) was separately dissolved in 2 ml DMF. The amine/NMM solution was introduced to the resin, followed by introduction of the PyAOP solution. The resin was incubated overnight at RT, followed by rinsing with MeOH (3 ml, 3min, 3x). The peptide was cleaved with 55 mmole of $Cu(OAc)_2$, and the MeCN and Pyr were removed by rotary evaporation. The remaining aqueous solution was frozen at -78 °C and lyophilized overnight. The solid was dissolved in 1.5 ml of 1/1 MeCN/H_2O (v/v) and purified by prep HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation, and aqueous remnants were frozen at -78 °C and lyophilized overnight. Purified yield: (1.4 mmole) 2%. High-res MS: found m/z 910.45740, calcd 910.45700 (M + H)$^{+}$; found m/z 908.44300, calcd 908.44240 (M - H)$^{-}$.

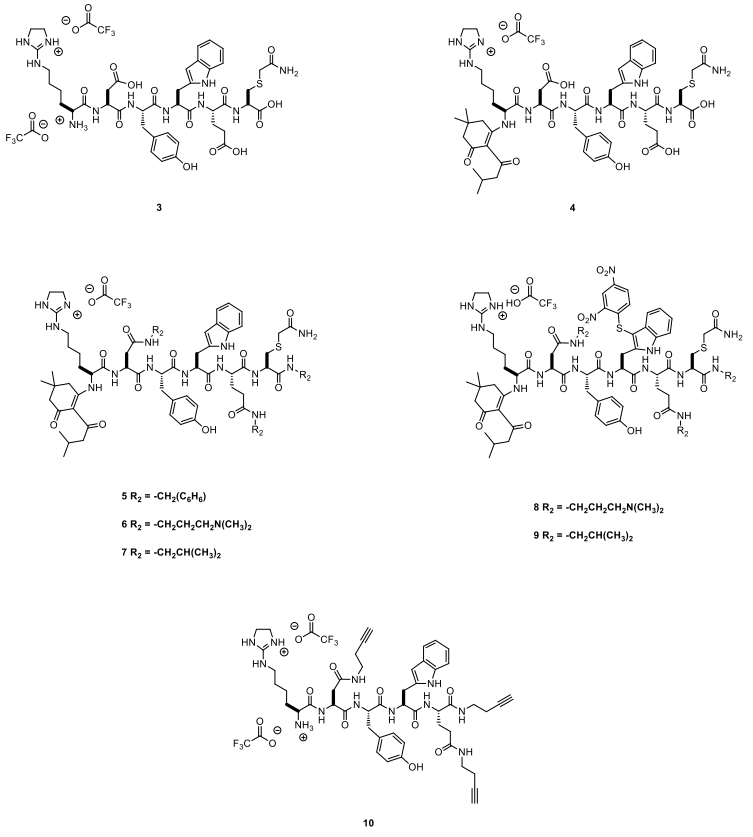
Tryptophan modification of immobilized peptide (11). Immobilized peptide 10 was prepared as described using 193 mg of the same resin. 2,4-Dinitrobenzenesulfenyl chloride (0.30 mmole) was dissolved in 3 ml of glacial acetic acid. This solution was introduced to the swollen resin, and incubated for 4 h at RT. The solution was removed from the resin and 6 ml of DMF was passed through the resin. Cleavage of peptide 11 from hydrazinobenzoyl resin using H_2O. Cleavage of the peptide was performed as described with copper acetate (0.3 mmole) dissolved in 3 ml of 45/45/10 MeCN/ H_2O/Pyr (v/v/v). MeCN and pyr were removed by rotary evaporation, and the remaining aqueous solution was frozen at -78 °C and lyophilized overnight. The solid was dissolved in 1.5 ml of 1/1/ MeCN/H2O (v/v) and purified by prep HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation. Aqueous remnants were frozen at
-78 °C and lyophilized overnight. Purified yield: (5.4 mmole) 4%. High-res MS: found m/z 1108.42840, calcd 1108.43050 $(M + H)^+$; found m/z 1106.41400, calcd 1106.41600 $(M - H)^-$.

Cleavage of peptide 12 from hydrazinobenzoyl resin using 1-amino-3-butyne. Copper acetate (0.33 mmole) was dissolved in 3 ml of 9/8.3/1.6 MeCN/Pyr/1-amino-3-butyne (v/v/v). Subsequently, the solution was introduced to swollen resin. The resin was incubated for 4 h at RT, followed by filtration to collect the solution, and the MeCN and pyridine were removed by rotary evaporation. Collected washes of the resin with DMF (3 ml, 3 min, 3x) were used to improve the solubility of the peptide. The solvent was removed by centrifugal evaporation (35 °C, 24 h). The solid was dissolved in 1.5 ml of 1/1 MeCN/ H_2O (v/v) and purified by prep HPLC. The organic solvent in the peptide fraction was removed via rotary evaporation and the aqueous remnants were frozen at -78 °C and lyophilized overnight. Purified yield: (5 mmole) 5%. High-res MS: found m/z 1159.47250, calcd 1159.47780 $(M + H)^+$; found m/z 1157.46220, calcd 1157.46330 $(M - H)^-$.

2.5 Acknowledgements

We gratefully acknowledge our funding sources: the National Institute of Health (5DP1GM106408), the Welch Foundation (F1515 to E. M. M.), the Defense Advanced Research Projects Agency (DARPA, N66001-14-2-4051), and the Welch Regents Chair to EVA (F-0046).

2.6 Experimental Characterization



(SI-Figure 1). KDYWEC and KDYWE derivatives.

Figure 2.4: Solid-phase intermediates.

Preparation of Modifying Reagents

2-Methoxy-4,5-dihydro-1H-imidazole was prepared following a literature protocol. (Peters EC, Horn DM, Tully DC, Brock A. A novel multifunctional modifying reagent for

enhanced protein characterization with mass spectrometry. *Rapid Commun. Mass Spectrom*. 2001; **15**: 2387-2392.)

1-(4,4-dimethyl-2,6-dioxocyclohexylidene)-3-methylbutyl diethyl phosphate was prepared by dissolving of 2-(3-methylbutyryl)-5,5-dimethyl-1,3-cyclohexandione (17 μmole) in 0.5 ml of dry MeCN under argon. Solution was placed in ice bath. DIPEA (20 μmole) was introduced, followed by slow introduction of diethylchlorophosphate (22 μmole). Reaction was stirred overnight at RT. Yield: quantitative. 1H NMR (500 MHz, Acetonitrile-d3) δ 4.12 (dq, J = 8.3, 7.1 Hz, 4H), 2.67 (d, J = 1.6 Hz, 2H), 2.45 (d, J = 6.8 Hz, 2H), 2.26 (s, 2H), 2.09 – 2.00 (m, 1H), 1.30 – 1.27 (m, 6H), 1.06 (s, 6H), 0.89 (d, J = 6.7 Hz, 6H). 13C NMR (126 MHz, cd3cn) δ 197.70, 164.75, 66.22, 66.17, 53.74, 51.03, 42.23, 33.22, 28.06, 24.81, 22.92, 16.42, 16.37. HR-res MS: found m/z 383.15970, calcd. 383.15940 $(M+Na^+)^+$; found m/z 359.16290, calcd. 359.16290 $(M-H)^-$ (Adapted from Zhang, H. A process for the preparation of the intermediate of β-methyl carbapenem. WO 2007104219 A1, September 20, 2007.)

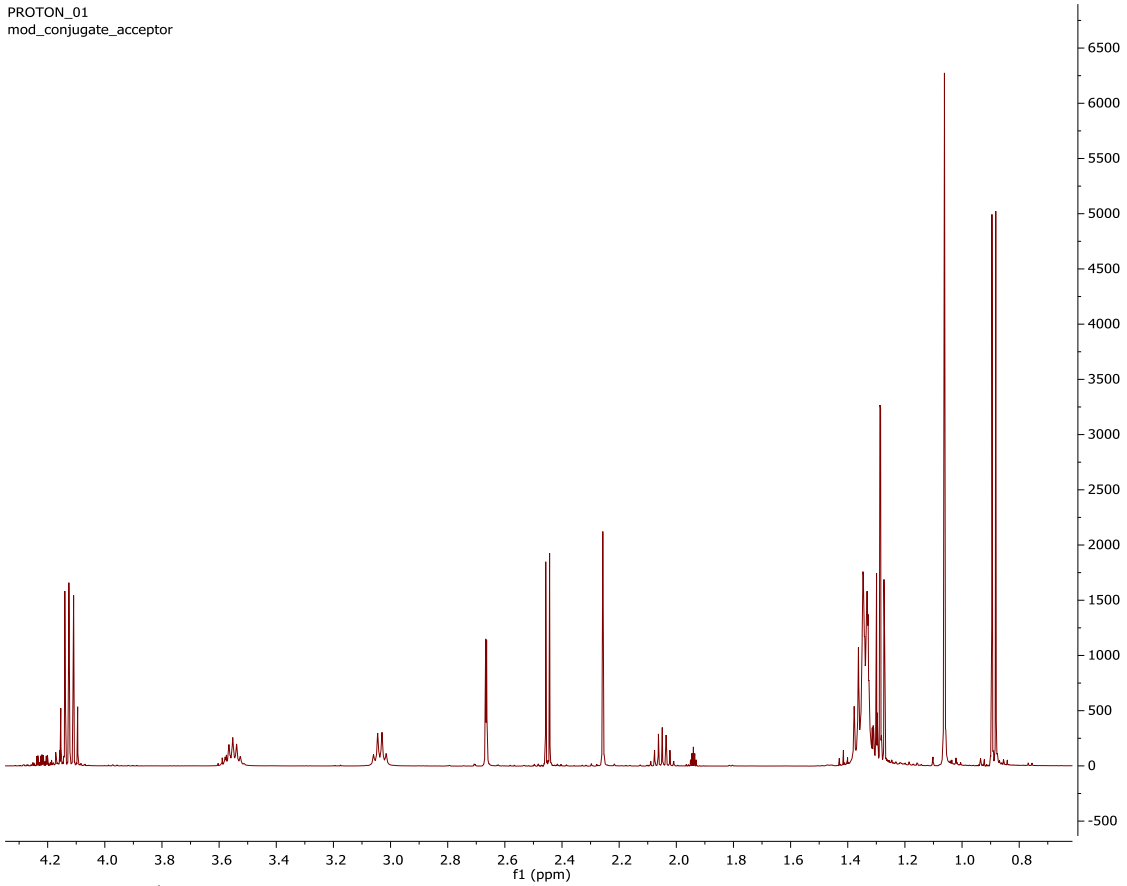


Figure 2.5: [1]Proton NMR of 5,5-dimethyl-2-(3-methylbutanoyl)-3-oxocyclohex-1-en-1-yl diethyl phosphate (Phos-DOD).

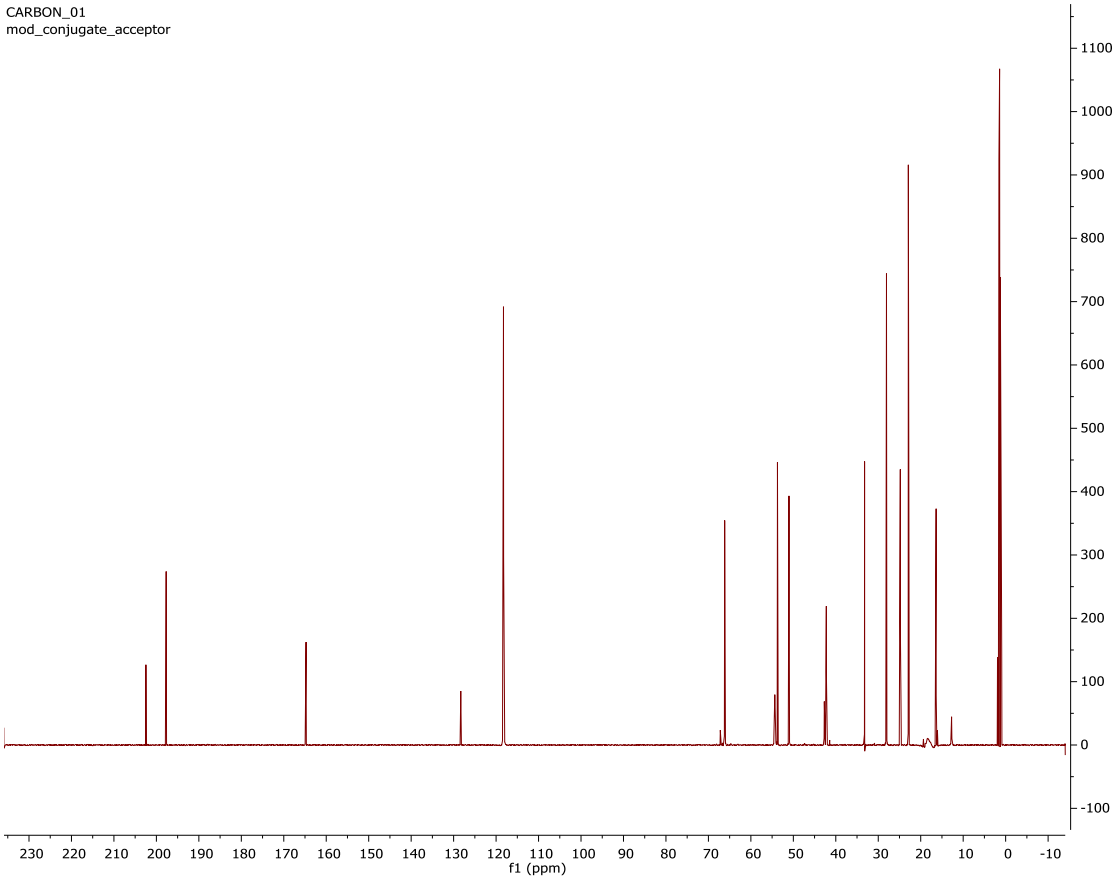


Figure 2.6: ^{13}Carbon NMR of phos-DOD.

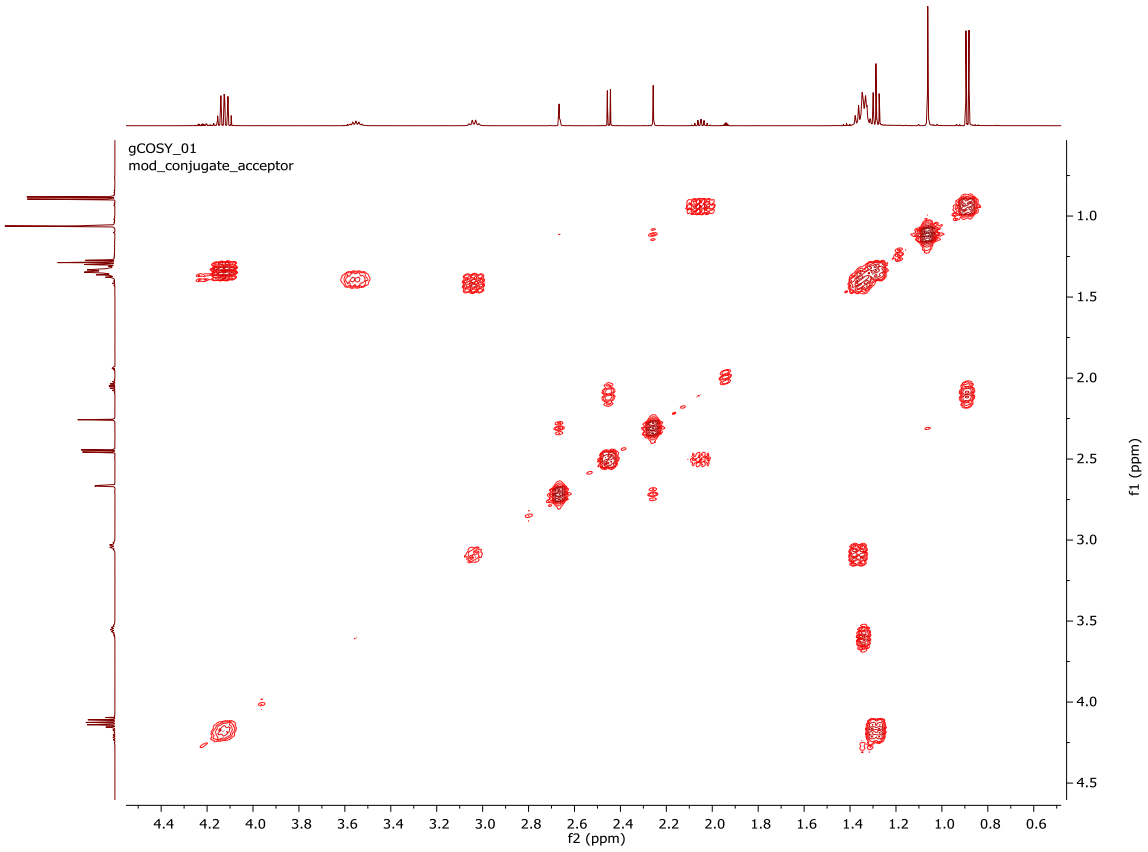
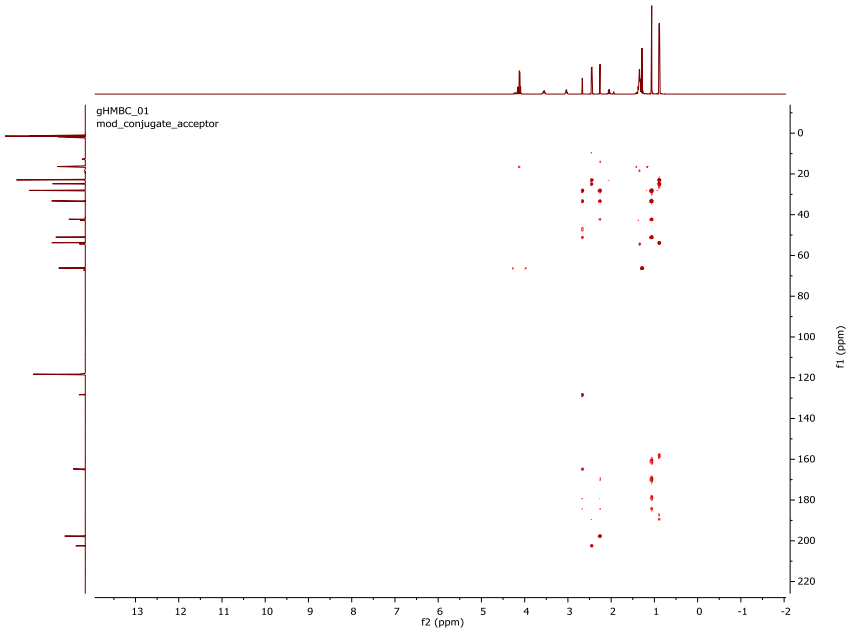


Figure 2.7: COSY NMR of phos-DOD.

(a)



(b)

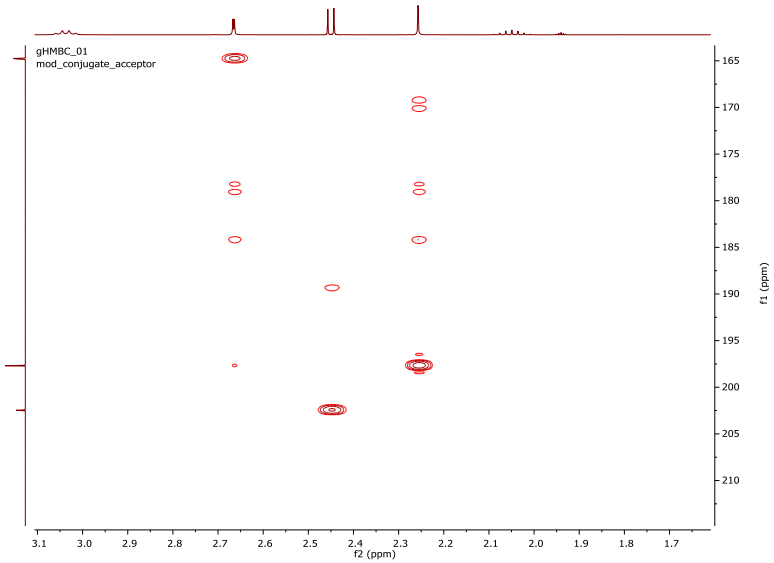


Figure 2.8: 2-D NMR for phos-DOD. 2-D NMR for phos-DOD.(a) HMBC NMR of phos-DOD. (b) Region shown demonstrated doublet attributed to position **a** in phos-DOD correlated with a carbon in the carbonyl region, instead of an alkene as shown in **1**.

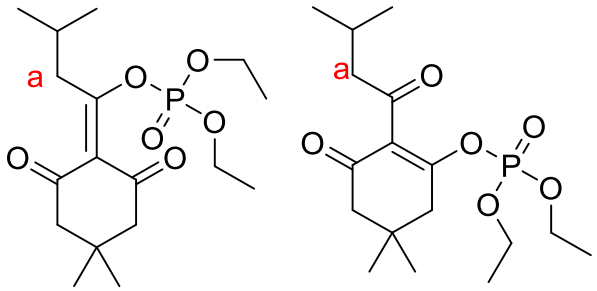
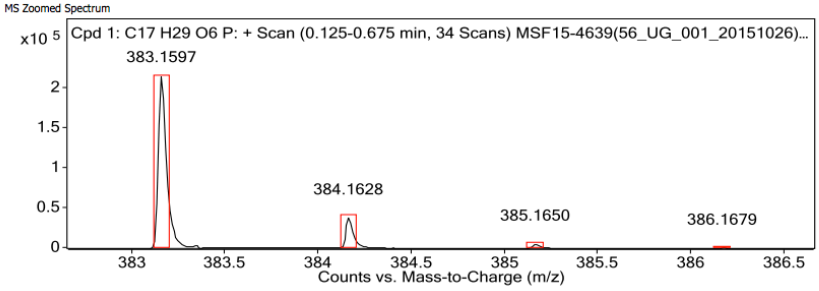


Figure 2.8 (cont.): (c) Compounds 1 and phos-DOD.

Target Compound Screening Report

Data File	MSF15-4639(56_UG_001_20151026)_hrESIpos1.d	**Sample Name**	4639(56_UG_001_20151026)	**Comment**	4639(56_UG_001_20151026)
Position	P1-B4	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	10/27/2015 2:34:51 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

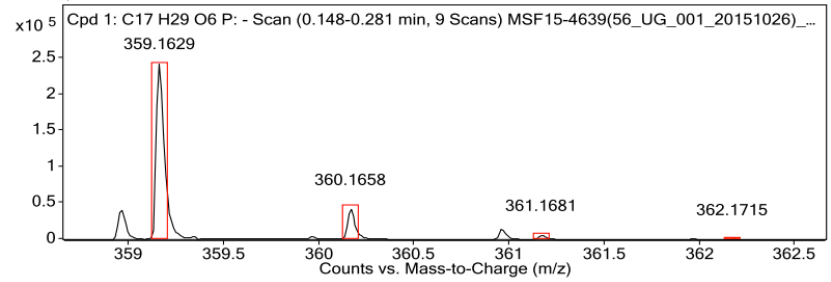
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
130.15910			2418499.71			
383.15970	383.15940	1	216028.95	C17H29O6P	(M+Na)+	-0.88
384.16280	384.16280	1	38420.76	C17H29O6P	(M+Na)+	-0.04
385.16500	385.16510	1	5741.21	C17H29O6P	(M+Na)+	0.45
386.16790	386.16780	1	640.61	C17H29O6P	(M+Na)+	-0.22

--- End Of Report ---

Target Compound Screening Report

Data File	MSF15-4639(56_UG_001_20151026)_hrESIneg1.d	**Sample Name**	4639(56_UG_001_20151026)	**Comment**	4639(56_UG_001_20151026)
Position	P1-B4	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	10/27/2015 2:49:13 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
331.13190			613191.49			
359.16290	359.16290	1	242412.53	C17H29O6P	(M-H)-	0.09
360.16580	360.16630	1	42336.24	C17H29O6P	(M-H)-	1.43
361.16810	361.16860	1	6543.88	C17H29O6P	(M-H)-	1.49
362.17150	362.17130	1	676.43	C17H29O6P	(M-H)-	-0.48

--- End Of Report ---

Figure 2.9: High-resolution mass spectrometry data for phos-DOD.

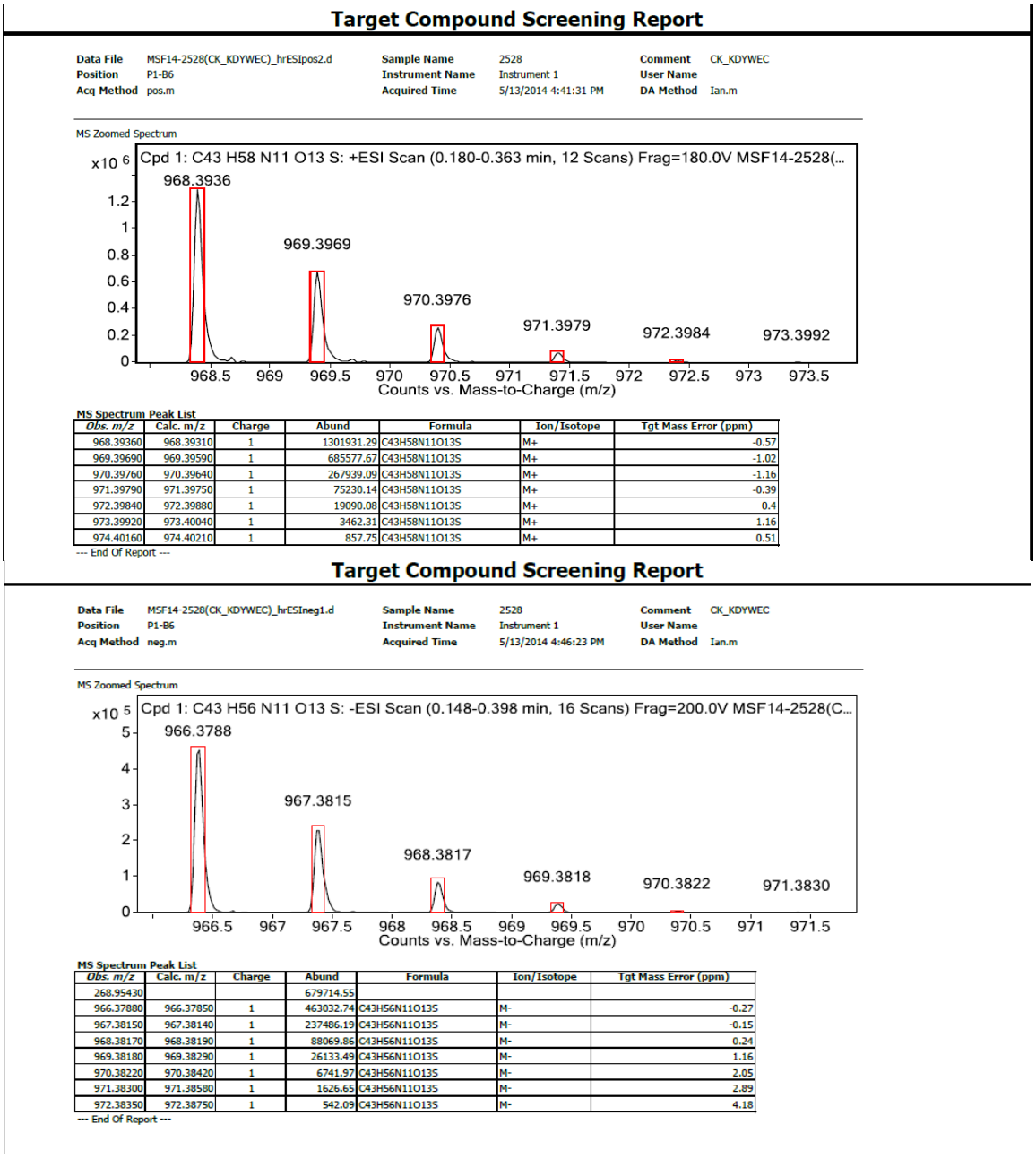


Figure 2.10: High-resolution mass spectrometry data for peptide 3.

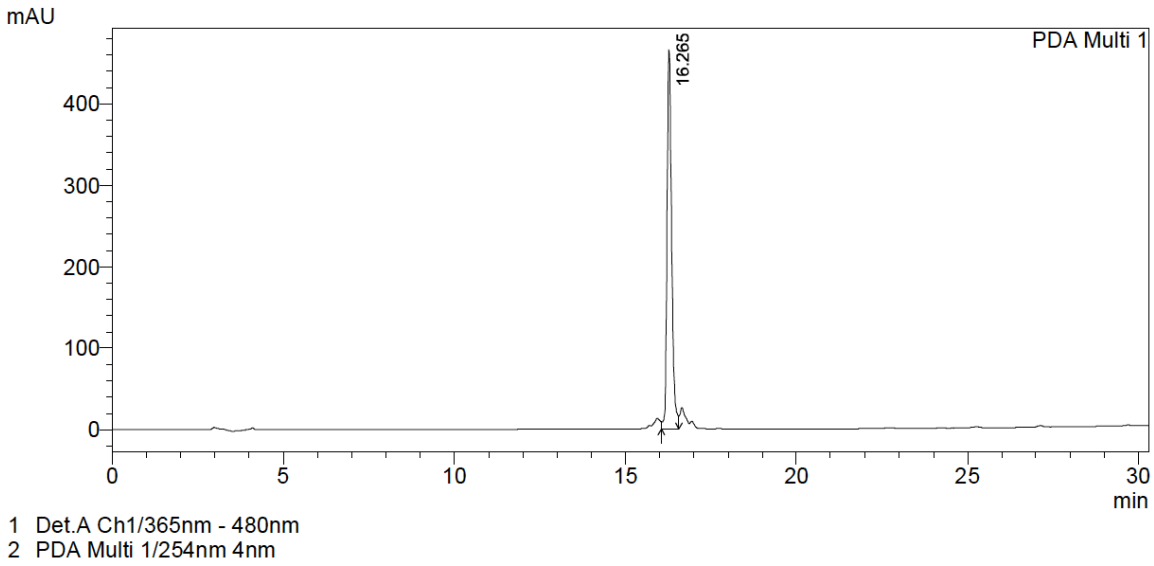


Figure 2.11.: HPLC trace for purified peptide 3.

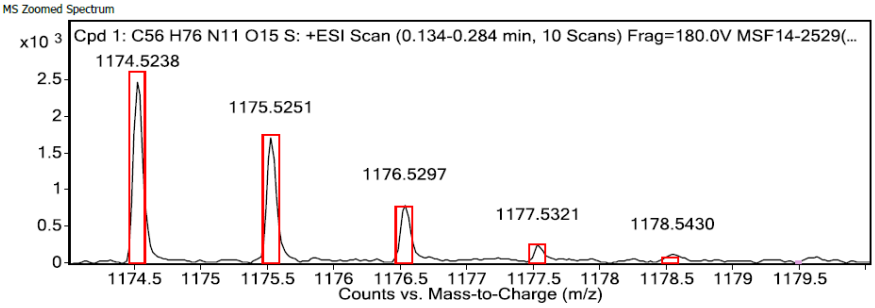
Desalting of peptide 4

Crude peptide was prepared for preparative HPLC using an Extract Clean^{TM} C_{18} 500 mg /4 ml solid phase extraction column. Column was flushed with 6 ml of 90/10 MeOH/H_2O with 0.1% TFA (v/v/v) at a flow rate of 1 drop sec^{-1} (RT), followed by equilibration with 3 ml of 0.1% TFA in water (v/v) at a flow rate of 1 drop sec^{-1}. Acidified peptide solution was loaded on the column 1 drop sec^{-1} (RT). Peptide was eluted with 1 ml 5% MeOH/Water with 0.1% TFA (v/v/v). Residually bound peptide was eluted with 50/50 MeCN/Water with 0.1% TFA (v/v/v).

Target Compound Screening Report

Data File	MSF14-2529(nCK_KDYWEC)_hrESIpos1.d	**Sample Name**	2529	**Comment**	nCK_KDYWEC
Position	P1-B7	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	5/13/2014 5:03:41 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

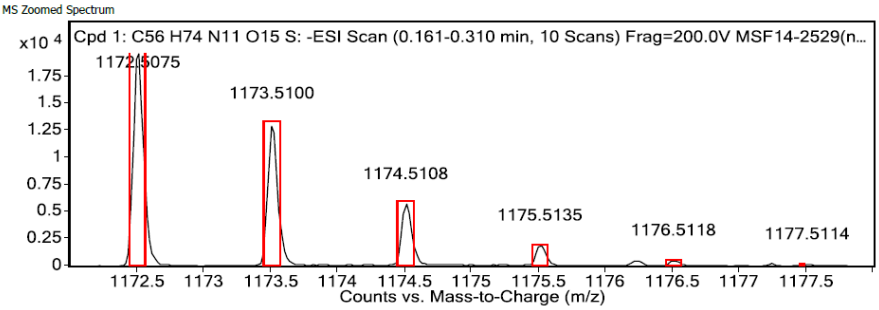
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
313.05800			819276.72			
1174.52380	1174.52380	1	2515.13	C56H76N11O15S	M+	-0.01
1175.52510	1175.52670	1	1723.84	C56H76N11O15S	M+	1.4
1176.52970	1176.52800	1	810.08	C56H76N11O15S	M+	-1.48
1177.53210	1177.52930	1	264.74	C56H76N11O15S	M+	-2.4
1178.54300	1178.53070	1	131.04	C56H76N11O15S	M+	-10.46

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-2529(nCK_KDYWEC)_hrESIneg1.d	**Sample Name**	2529	**Comment**	nCK_KDYWEC
Position	P1-B7	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	5/13/2014 5:10:44 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
585.75120			654506.35			
1172.50750	1172.50920	1	20019.71	C56H74N11O15S	M-	1.46
1173.51000	1173.51220	1	13205.15	C56H74N11O15S	M-	1.84
1174.51080	1174.51340	1	5758.11	C56H74N11O15S	M-	2.2
1175.51350	1175.51470	1	1930.16	C56H74N11O15S	M-	1.04
1176.51180	1176.51610	1	546.85	C56H74N11O15S	M-	3.69
1177.51140	1177.51780	1	198.53	C56H74N11O15S	M-	5.43

--- End Of Report ---

Figure 2.12:. High-resolution mass spectrometry data for peptide 4.

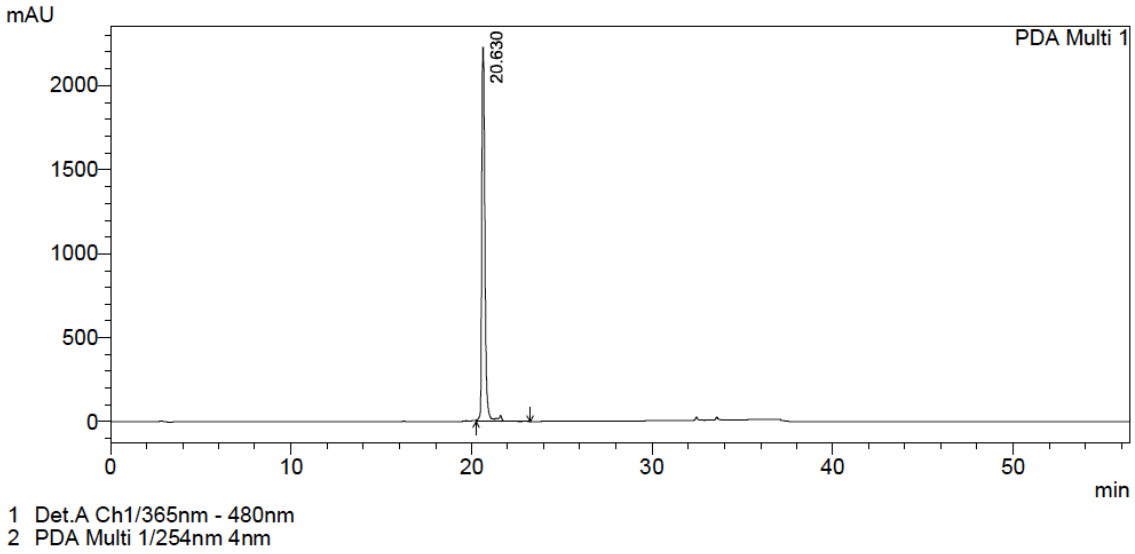
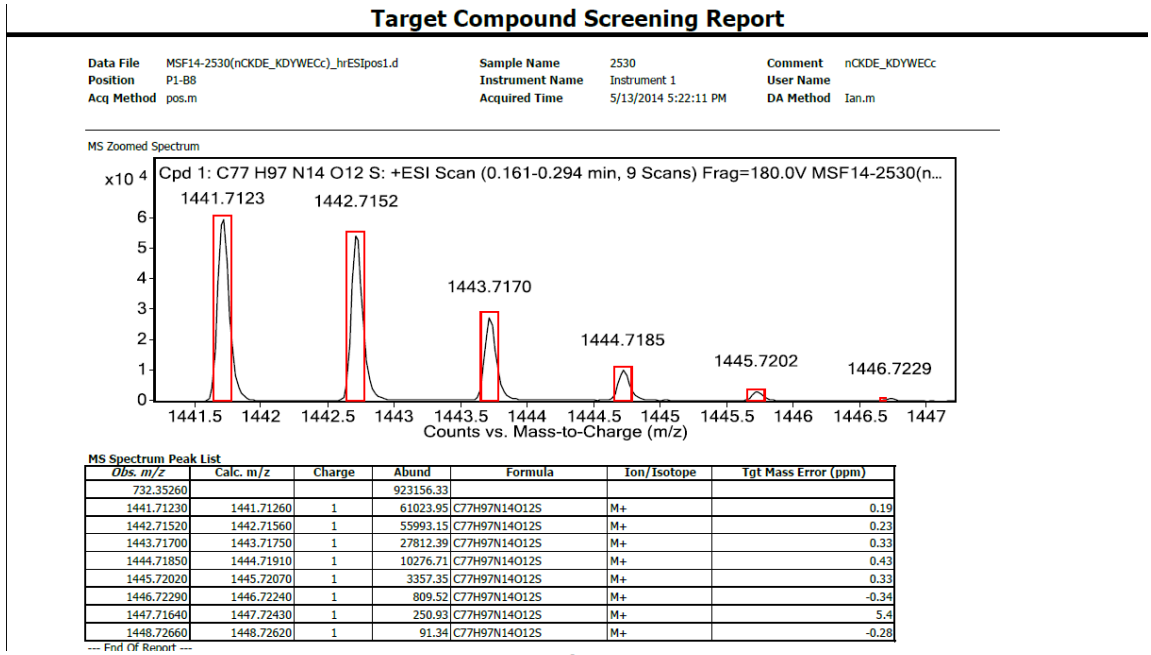


Figure 2.13: HPLC trace for purified peptide 4.



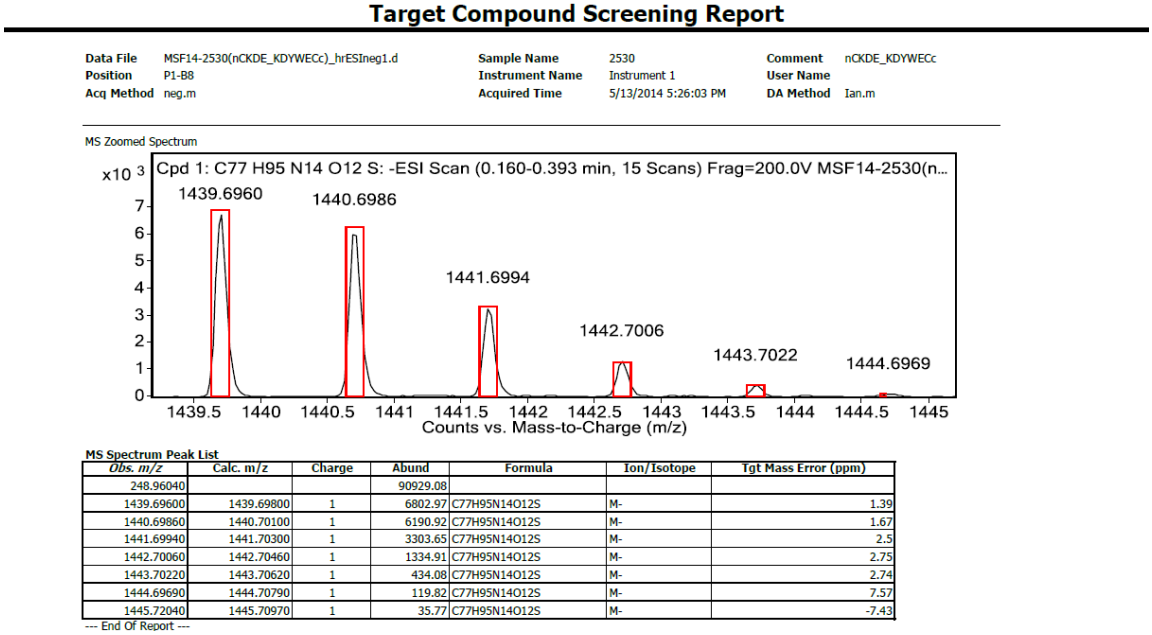


Figure 2.14: High-resolution mass spectrometry data for peptide 5.

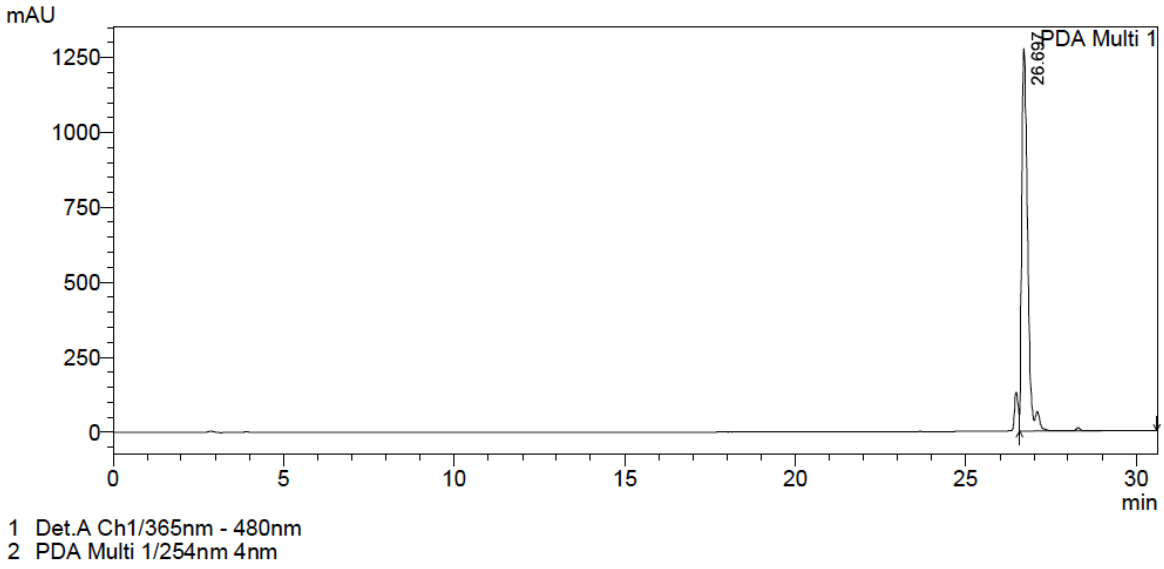
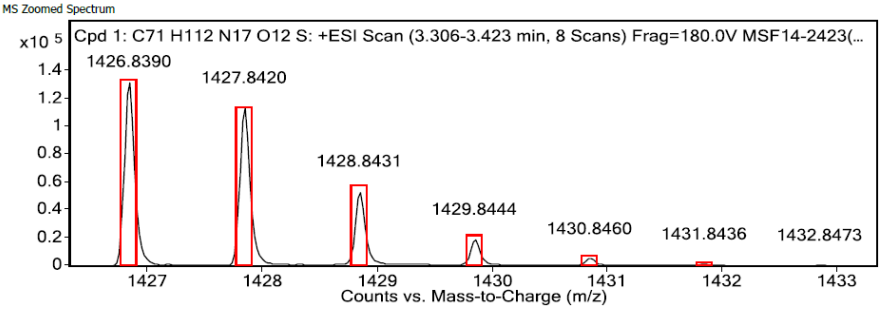


Figure 2.15: HPLC trace for purified peptide 5.

Target Compound Screening Report

Data File	MSF14-2423(nCKDE_KDYWECc_b)_hrESIpos2.d	**Sample Name**	2423(nCKDE_KDYWECc_b)	**Comment**	2423(nCKDE_KDYWECc_b)
Position	P1-F1	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos_column_20min.m	**Acquired Time**	6/25/2014 11:55:56 AM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
357.46650			2350276.17			
1426.83900	1426.83920	1	132784.34	C71H112N17O12S	M+	0.1
1427.84200	1427.84210	1	113609.25	C71H112N17O12S	M+	0.05
1428.84310	1428.84380	1	53501.04	C71H112N17O12S	M+	0.53
1429.84440	1429.84530	1	18949.68	C71H112N17O12S	M+	0.63
1430.84600	1430.84680	1	4918.58	C71H112N17O12S	M+	0.58
1431.84360	1431.84840	1	1160.52	C71H112N17O12S	M+	3.34
1432.84730	1432.85020	1	271.57	C71H112N17O12S	M+	2
1433.85390	1433.85210	1	86.89	C71H112N17O12S	M+	-1.24

--- End Of Report ---

Figure 2.16: High-resolution mass spectrometry data for peptide 6.

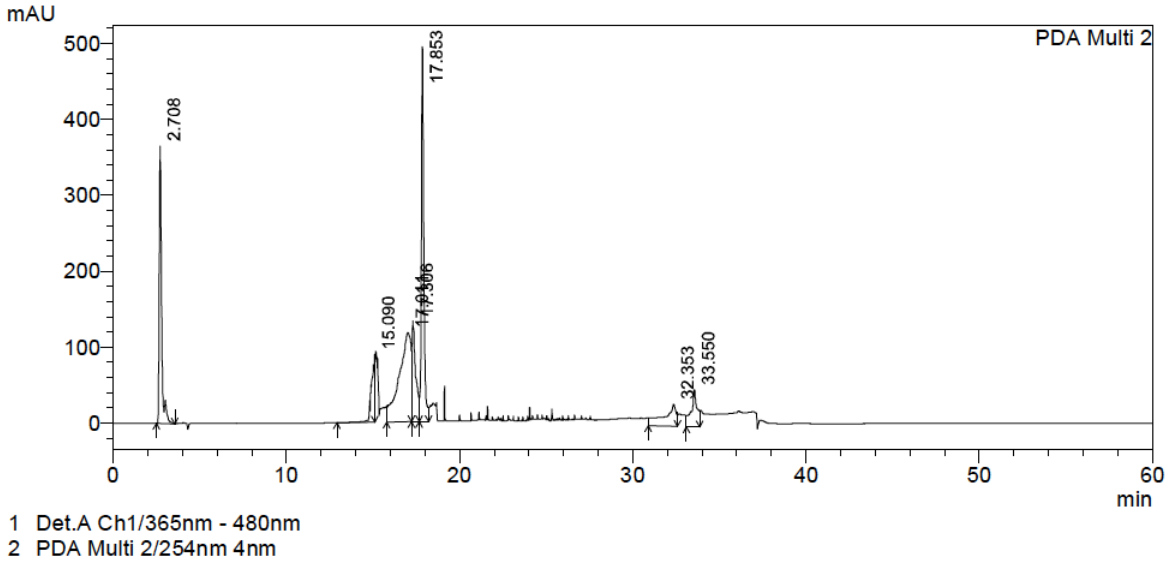
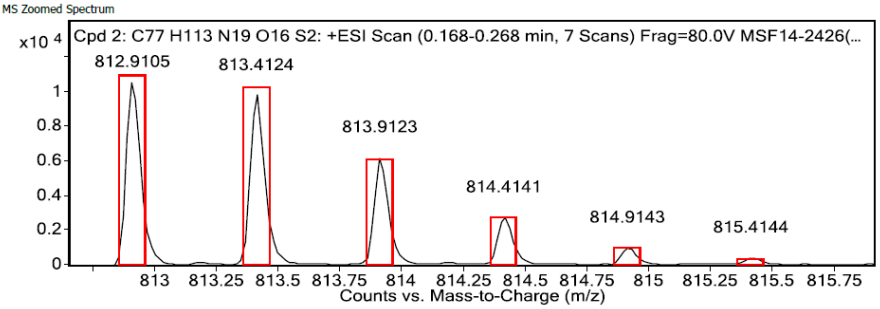


Figure 2.17: HPLC trace for purified peptide 6.

Target Compound Screening Report

Data File	MSF14-2426(nCKDEW_KDYWECc)_hrESIpos3.d	**Sample Name**	2426(nCKDEW_KDYWECc)	**Comment**	2426(nCKDEW_KDYWECc)
Position	P1-E1	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	9/11/2014 4:06:51 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

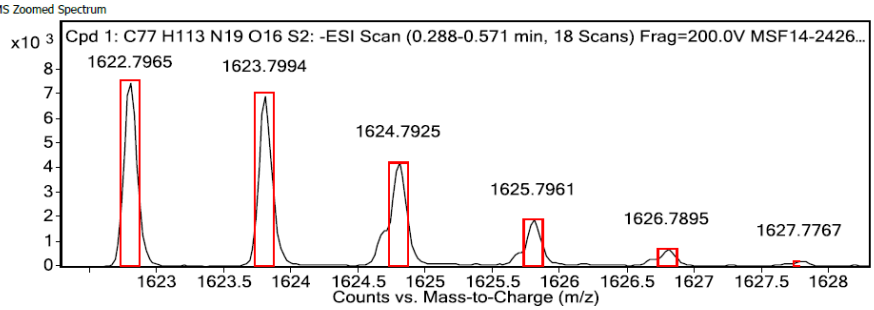
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
342.27890			487108.04			
812.91050	812.91000	2	10702.85	C77H113N19O16S2	(M+2H)+2	-0.65
813.41240	813.41140	2	9851.45	C77H113N19O16S2	(M+2H)+2	-1.2
813.91230	813.91200	2	6261.34	C77H113N19O16S2	(M+2H)+2	-0.39
814.41410	814.41250	2	2794.08	C77H113N19O16S2	(M+2H)+2	-1.9
814.91430	814.91300	2	1043.97	C77H113N19O16S2	(M+2H)+2	-1.58
815.41440	815.41350	2	468.78	C77H113N19O16S2	(M+2H)+2	-1.06

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-2426(nCKDEW_KDYWECc)_hrESIneg1.d	**Sample Name**	2426(nCKDEW_KDYWECc)	**Comment**	2426(nCKDEW_KDYWECc)
Position	P1-E1	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	9/11/2014 4:12:30 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
248.96060			234543.08			
1622.79650	1622.79810	1	7509.6	C77H113N19O16S2	(M-H)-	1.01
1623.79940	1623.80100	1	6927.6	C77H113N19O16S2	(M-H)-	0.96
1624.79250	1624.80220	1	4266.76	C77H113N19O16S2	(M-H)-	5.93
1625.79610	1625.80320	1	1905.57	C77H113N19O16S2	(M-H)-	4.36
1626.78950	1626.80420	1	710.54	C77H113N19O16S2	(M-H)-	9.05
1627.77670	1627.80520	1	235.21	C77H113N19O16S2	(M-H)-	17.53

--- End Of Report ---

Figure 2.18: High-Resolution Mass Spectrometry Data for Peptide 8.

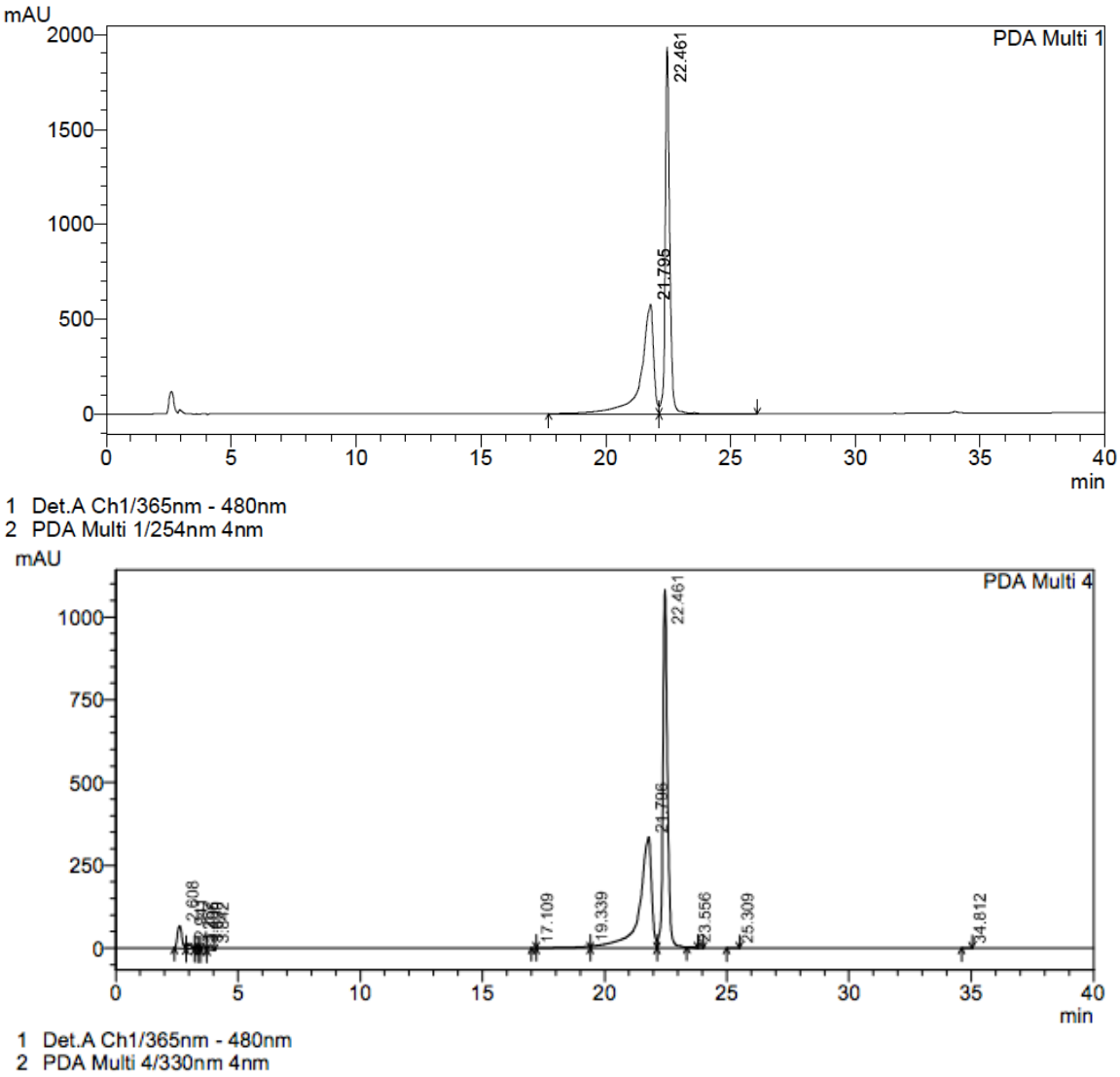
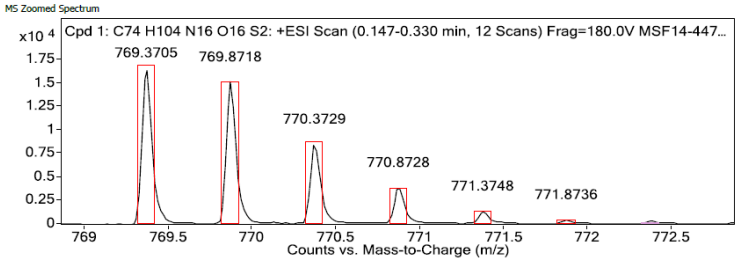


Figure 2.19: HPLC trace for purified peptide 8.

Target Compound Screening Report

Data File	MSF14-4477(nCKDEW_KDYWECc_isobutylamine)_hrESIpos1.d	**Sample Name**	4477(nCKDEW_KDYWECc_isobutylamine)	**Comment**	4477(nCKDEW_KDYWECc_isobutylamine)
Position	P1-E4	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	12/17/2014 1:17:46 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

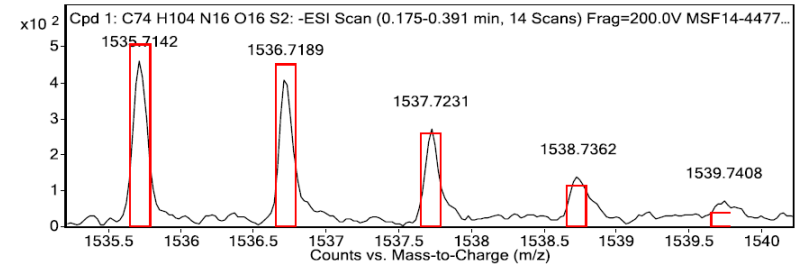
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
769.37050	769.37020	2	16614.5	C74H104N16O1652	(M+2H)+2	-0.49
769.87180	769.87160	2	15109.76	C74H104N16O1652	(M+2H)+2	-0.23
770.37290	770.37220	2	8503.04	C74H104N16O1652	(M+2H)+2	-0.97
770.87280	770.87260	2	3884.32	C74H104N16O1652	(M+2H)+2	-0.18
771.37480	771.37310	2	1382.74	C74H104N16O1652	(M+2H)+2	-2.14
771.87360	771.87360	2	462.97	C74H104N16O1652	(M+2H)+2	0.05
780.36040			37856.99			

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-4477(nCKDEW_KDYWECc_isobutylamine)_hrESIneg1.d	**Sample Name**	4477(nCKDEW_KDYWECc_isobutylamine)	**Comment**	4477(nCKDEW_KDYWECc_isobutylamine)
Position	P1-E4	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	12/17/2014 1:25:47 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
248.96000			64305.2			
1535.71420	1535.71850	1	463.88	C74H104N16O16S2	(M-H)-	2.82
1536.71890	1536.72140	1	417.47	C74H104N16O16S2	(M-H)-	1.61
1537.72310	1537.72250	1	273.98	C74H104N16O16S2	(M-H)-	-0.38
1538.73620	1538.72350	1	140.73	C74H104N16O16S2	(M-H)-	-8.25
1539.74080	1539.72440	1	74.22	C74H104N16O16S2	(M-H)-	-10.66

--- End Of Report ---

Figure 2.20: High resolution mass spectrometry data for peptide 9.

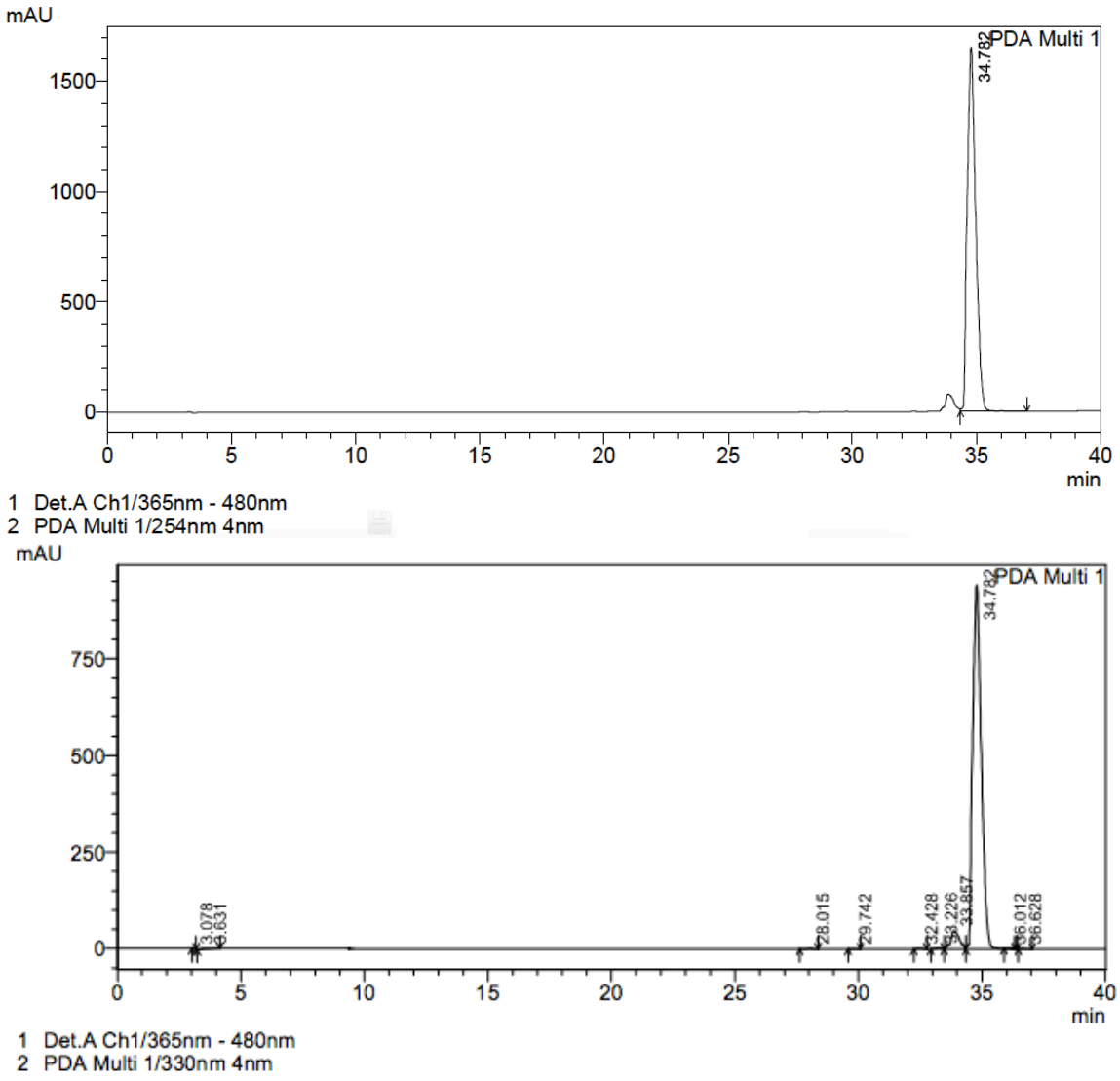
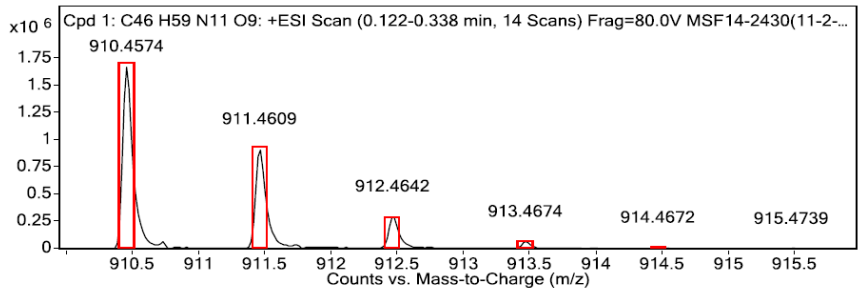


Figure 2.21: HPLC trace for purified peptide 9.

Target Compound Screening Report

Data File	MSF14-2430(11-2-14_KDE_KDYWE)_hrESIpos3.d	**Sample Name**	2430(11-2-14_KDE_KDYWE)	**Comment**	2430(11-2-14_KDE_KDYWE)
Position	P1-F6	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	11/13/2014 3:04:18 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

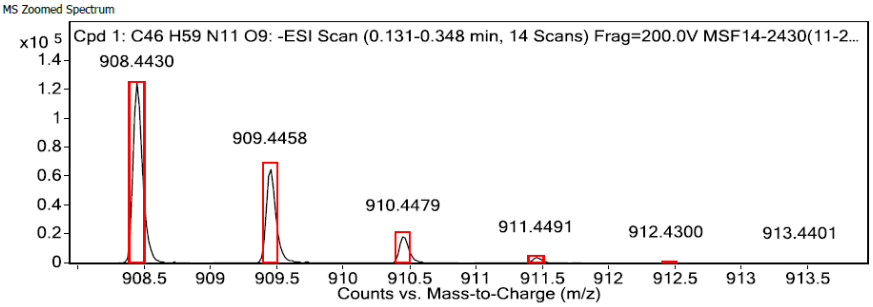
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
455.73310			2158894.04			
910.45740	910.45700	1	1675021.78	C46H59N11O9	(M+H)+	-0.41
911.46090	911.45990	1	921286.98	C46H59N11O9	(M+H)+	-1.04
912.46420	912.46270	1	301557.87	C46H59N11O9	(M+H)+	-1.68
913.46740	913.46530	1	68797.4	C46H59N11O9	(M+H)+	-2.31
914.46720	914.46790	1	14291.77	C46H59N11O9	(M+H)+	0.73
915.47390	915.47050	1	3706.55	C46H59N11O9	(M+H)+	-3.76

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-2430(11-2-14_KDE_KDYWE)_hrESIneg2.d	**Sample Name**	2430(11-2-14_KDE_KDYWE)	**Comment**	2430(11-2-14_KDE_KDYWE)
Position	P1-F6	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	11/13/2014 3:08:03 PM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
908.44300	908.44240	1	125450.88	C46H59N11O9	(M-H)-	-0.56
909.44580	909.44540	1	65993.42	C46H59N11O9	(M-H)-	-0.47
910.44790	910.44810	1	19498.3	C46H59N11O9	(M-H)-	0.23
911.44910	911.45080	1	4609.84	C46H59N11O9	(M-H)-	1.82
912.43000	912.45340	1	1120.5	C46H59N11O9	(M-H)-	25.63
913.44010	913.45590	1	261.95	C46H59N11O9	(M-H)-	17.34
1022.43610			159301.06			

--- End Of Report ---

Figure 2.22: High resolution mass spectrometry data for peptide 10.

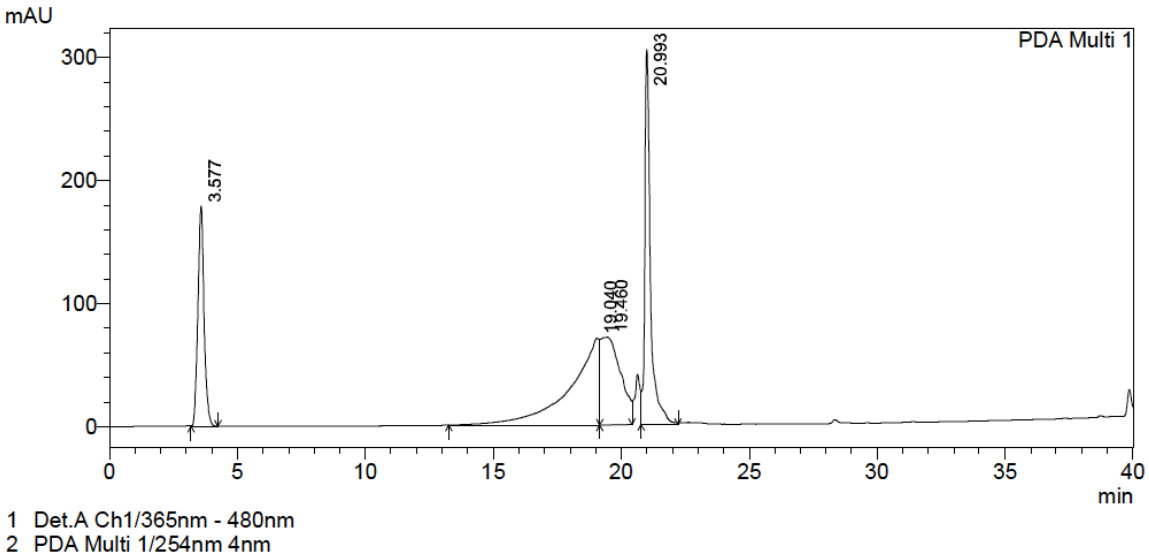
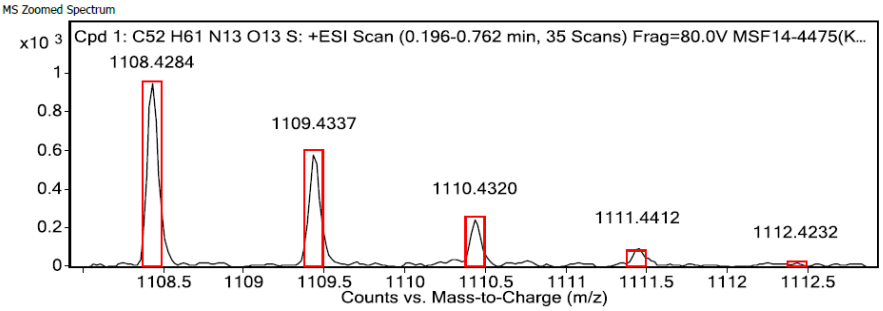


Figure 2.23: HPLC trace for purified peptide 10.

Target Compound Screening Report

Data File	MSF14-4475(KDEW_KDYWE)_hrESIpos2.d	**Sample Name**	4475(KDEW_KDYWE)	**Comment**	4475(KDEW_KDYWE)
Position	P1-B5	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	11/18/2014 9:26:15 AM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

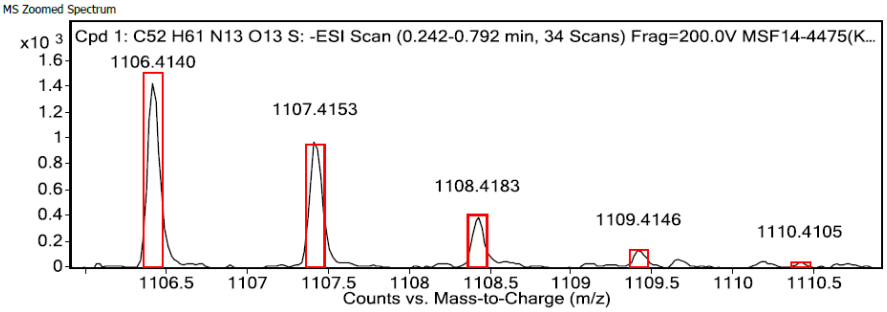
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
554.71870			6263.03			
1108.42840	1108.43050	1	953.45	C52H61N13O13S	(M+H)+	1.94
1109.43370	1109.43340	1	586.36	C52H61N13O13S	(M+H)+	-0.24
1110.43200	1110.43440	1	248.59	C52H61N13O13S	(M+H)+	2.2
1111.44120	1111.43560	1	97.3	C52H61N13O13S	(M+H)+	-5.1
1112.42320	1112.43690	1	21.73	C52H61N13O13S	(M+H)+	12.36

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-4475(KDEW_KDYWE)_hrESIneg2.d	**Sample Name**	4475(KDEW_KDYWE)	**Comment**	4475(KDEW_KDYWE)
Position	P1-B5	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	11/18/2014 9:23:11 AM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
979.96560			1505.41			
1106.41400	1106.41600	1	1434.94	C52H61N13O13S	(M-H)-	1.76
1107.41530	1107.41880	1	991.33	C52H61N13O13S	(M-H)-	3.17
1108.41830	1108.41990	1	396.96	C52H61N13O13S	(M-H)-	1.44
1109.41460	1109.42100	1	135.62	C52H61N13O13S	(M-H)-	5.74
1110.41050	1110.42240	1	42.51	C52H61N13O13S	(M-H)-	10.67

--- End Of Report ---

Figure 2.24: High-resolution mass spectrometry data for peptide 11.

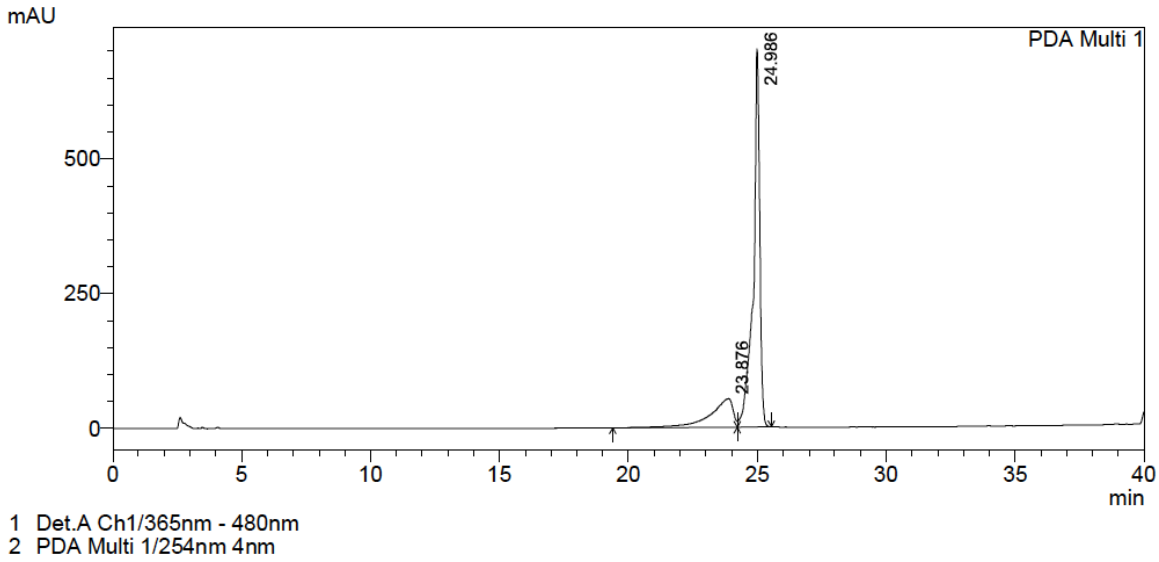
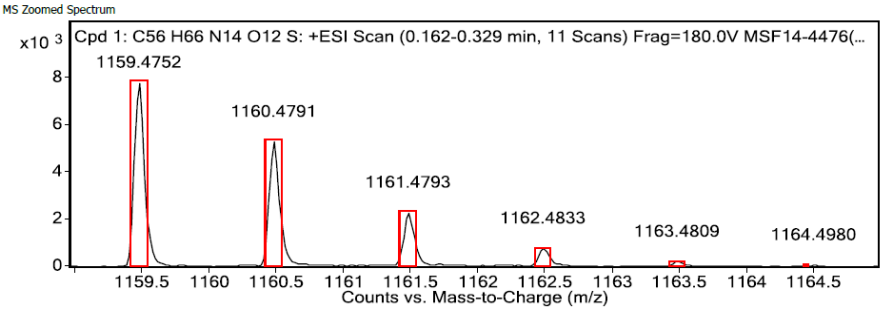


Figure 2.25: HPLC trace for purified peptide 11.

Target Compound Screening Report

Data File	MSF14-4476(KDEW_KDYWEc)_hrESIpos1.d	**Sample Name**	4476(KDEW_KDYWEc)	**Comment**	4476(KDEW_KDYWEc)
Position	P1-B5	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	pos.m	**Acquired Time**	12/16/2014 8:02:31 AM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

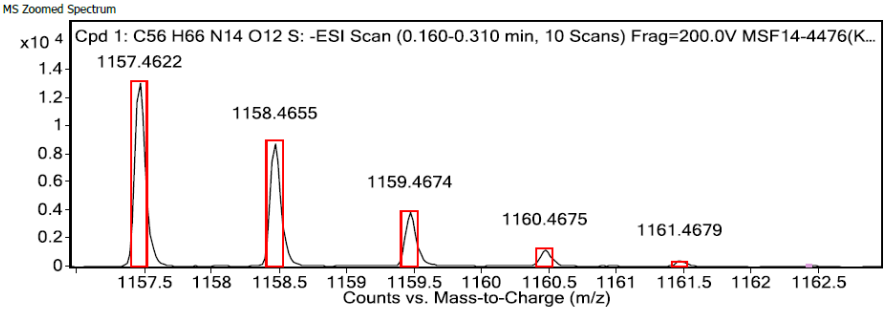
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
580.24270			21918.24			
1159.47520	1159.47780	1	7795.29	C56H66N14O12S	(M+H)+	2.24
1160.47910	1160.48070	1	5294.49	C56H66N14O12S	(M+H)+	1.38
1161.47930	1161.48190	1	2299.1	C56H66N14O12S	(M+H)+	2.27
1162.48330	1162.48310	1	780.91	C56H66N14O12S	(M+H)+	-0.18
1163.48090	1163.48440	1	221.77	C56H66N14O12S	(M+H)+	3.06
1164.49800	1164.48600	1	78.18	C56H66N14O12S	(M+H)+	-10.3

--- End Of Report ---

Target Compound Screening Report

Data File	MSF14-4476(KDEW_KDYWEc)_hrESIneg1.d	**Sample Name**	4476(KDEW_KDYWEc)	**Comment**	4476(KDEW_KDYWEc)
Position	P1-B5	**Instrument Name**	Instrument 1	**User Name**	
Acq Method	neg.m	**Acquired Time**	12/16/2014 8:05:08 AM	**DA Method**	Ian.m

MS Zoomed Spectrum



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
112.98560			19732.4			
1157.46220	1157.46330	1	13140.28	C56H66N14O12S	(M-H)-	0.9
1158.46550	1158.46610	1	8800.68	C56H66N14O12S	(M-H)-	0.56
1159.46740	1159.46740	1	3897.37	C56H66N14O12S	(M-H)-	0
1160.46750	1160.46850	1	1237.52	C56H66N14O12S	(M-H)-	0.9
1161.46790	1161.46990	1	464.1	C56H66N14O12S	(M-H)-	1.66

--- End Of Report ---

Figure 2.26: High-resolution mass spectrometry data for peptide 12.

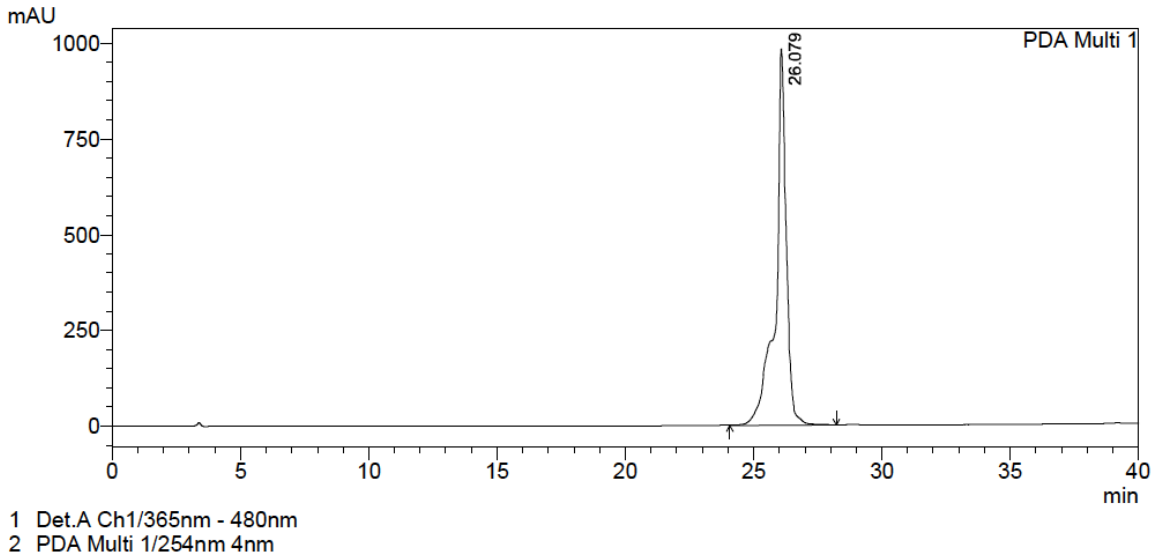


Figure 2.27: HPLC trace for purified peptide 12.

2.7 References

1. Swaminathan, A. A. Boulgakov, E. M. Marcotte, *PLoS Comp. Bio.*, 2015, **11(2)**, 1-17.
2. Y. Yao, M. Docter, J. V. Ginkel, and C. Joo, Phys. Biol., 2015., **12,** 1-6.
3. Julka, and F. Regnier, *J. Proteome Res.*, 2004, **3**, 350-363.
4. S. L. Cockrill, K. L. Foster, J. Wildsmith, A. R. Goodrich, J. G. Dapron, T. C. Hassel, W.K. Kappel, and G. B. I. Scott, *Biotech.*, 2005, **38**, 301-304.
5. B. L. Frey, D. T. Ladror, S. B. Sondalle, C.J. Krusemark, A. L. Jue, J. J. Coon, and L. M. Smith, *J. Am. Mass Spectrom*, 2013, **24**, 1710-1721.
6. C.J . Krusemark, B. L. Frey, L. M. Smith, and P. J. Belshaw, *Gel-Free Proteomics, Methods in Molecular Biology*, **753** (Eds: Gevaert K, Vandekerckhove J) Humana Press, New York, 2011, pp.77-91.
7. H. R. Horton, and D. E. Koshland, *J. Am. Chem. Soc.* 1965, **87**, 1126-1132.
8. E. Scoffone, and A. Fontana, R. Rocchi, *Biochem*, 1968, **7**, 971-979.
9. H. Kuyama, M. Watanabe, C. Toda, E. Ando, K. Tanaka, and O. Nishimura, *Rapid Commun. Mass Spectrom.*, 2003, **17**, 1642-1650.
10. J. I. Macdonald, H. K. Munch, T. Moore, and M. B. Francis, *Nat. Chem. Bio.* 2015, **11**, 326-334.
11. M. Bantscheff, M. Bantscheff, G. Sweetman, and J. Rick, B. Kuster, *Anal. Bioanal. Chem.*, 2007, **389**, 1017-1031.
12. C. D. Spicer, and B. G. Davis, Nat. Comm. 2014, **5**, 1-14.
13. N. Krall, F. da Cruz, O. Boutureira, and G. J. L. Bernardes, *Nat. Chem.*, 2015, **8**, 1-11.
14. J. M. Chalker, and G. J. L. Bernardes, Y. A. Lin, B.G. Davis, *Chem. Asian J.* 2009, **4**, 630-640.
15. A. Isidro-Llobet, M. Alvarez, and F. Albericio, *Chem. Rev.*, 2009, **109**, 2455-2504.

16. B. J. Ko, and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 1991-2000.
17. J. R.Moffet, and M. A.Namboodiri, *Immunol. Cell Biol.* 2003, **81**, 247-265.
18. E. Scoffone, A. Fontana, and R. Rocchi, *Biochem. Biophys. Res. Commun.*, 1966, **25**, 170-174.
19. V. Wittman, and S. Seeberger, *Angew. Chem. Int. Ed.*, 2000, **39**, 4348-4352.
20. J. Tulla-Puche, and F. Albericio, The (classic concept of) solid support. In *The power of functional resins in organic synthesis* (Eds: Tulla-Pucha J, Albericio F) Wiley, Weinheim, 2008, pp. 3-14.
21. C. R. Millington, R. Quarell, and G. Lowe, *Tett. Lett.*, 1998, **39**, 7201-7204.
22. C. Rosenbaum, and H. Waldmann, *Tett. Lett.* 2001, **42**, 5677-5680.
23. T. Keough, M. P. Lacey, and R. S. Yongquist, *Rapid Commun. Mass Spectrom.*, 2000, **14**, 2348-2356.
24. W. Tang, and M. L. Becker, *Chem. Soc. Rev.* 2014, **43**, 7013-7059.
25. L. Zervas, D. Borovas, and E. Gazis, *J. Am. Chem. Soc.* 1963, **85**: 3660-3666.

Chapter 3: Fluorescently Labeled Model Peptides for a New Single-Molecule Peptide Sequencing Platform Using Edman Degradation and Total Internal Reflection Fluorescence (TIRF) Microscopy[3]

3.1 INTRODUCTION

3.1.1 Mass Spectrometry: Current Peptide Sequencing Approach and Limitations

Proteomics focuses on studying the structure and function of proteins.[1, 2] Understanding the proteins present in a cell at biological states is a key goal in the field. Knowing which proteins are expressed or under expressed at certain conditions allows for a more detailed description of an organism. Current technologies, such as mass spectrometry, have revolutionized the field by increasing the level of sensitivity and throughput when analyzing complex mixtures of protein digests. However, there are still limitations with this technology. One important barrier is the difficulty in resolving lower abundant proteins from more prevalent ones. Several biomarkers for diseases like prostate cancer or Alzheimer's could be better detected if key proteins, which are found at smaller concentrations, can be identified.[3, 4]

Mass spectrometry uses reverse-phase column separation methods to increase resolution. Yet, it does not guarantee that peptide digests reach the detector a single compound at a time. Short peptides (15-20 residues) with similar hydrophobic/hydrophilic content usually elute at similar times.[5] For instance, if one lower-abundant peptide elutes closely to a more abundant one, the signal of the less prevalent peptide will be obscured. In addition, depending on the side chains present on a protein digest, some peptides will

[3] Swaminathan, Jagannath[†]; Boulgakov, Alexander A.[†]; **Hernandez, Erik T.**[†]; Bardo, Angela M.[†]; Bachman, Logan; Johnson, Amber; Marotta, Joseph; Anslyn, Eric V.; Marcotte, Edward M. "Peptide Sequence Information can be Obtained at the Level of a Single Molecule." *Manuscript submitted to Nature Biotechnology*. Prepared model peptides and surface chemistries.

inherently ionize more efficiently. A more efficient ionization of a peptide increases the signal intensity, and could exacerbate the ability to detect peptides if they lack residues that ionize.[6] For example, peptides rich in arginine, histidine, or lysine can be protonated forming a positive charge.[7] Peptides with these residues will be observed more readily in the positive mode than those without them. So, if a peptide at smaller concentrations that elutes closely with a more abundant peptide and if both these peptides do not ionize well, observing the less concentrated peptide becomes less likely. Efforts have been made to improve ionization of peptides by chemical modifications of proteins, enrichment protocols, and instrument tuning, but these approaches add to the time required for sample preparation and the amount of sample needed.[8]

3.1.2 Higher Sensitivity Required

A more effective technique is required for observing less abundant proteins. Although mass spectrometry can detect at picamole concentrations, issues that arise with complex mixtures and ionization efficiencies obstruct that sensitivity. Therefore, new sequencing platforms that can overcome these challenges are needed to improve upon the current effective sensitivity available. The Marcotte group at the University of Texas at Austin has been working on a new platform that seeks to solve current challenges in peptide/protein sequencing.[9]

3.1.3 Single-Molecule DNA Sequencing

Unlike next-generation DNA sequencing technologies, which increased the throughput and level of detection for identifying oligonucleotides, protein sequencing did not experience the same technology advancement. In the early 2000's, several technologies emerged aimed at improving from the Sanger method for sequencing DNA.[10, 11] One such method was based on based single-stranded DNA (ssDNA) and the detection of the complementary strand using Förster Resonance Energy Transfer (FRET).[12] Other

technologies took advantage of DNA polymerase and fluorescently labeled oligonucleotides. Total Internal Reflection Fluorescence (TIRF) microscopy was used for ssDNA immobilized on a glass surface. As the DNA polymerase incorporated a fluorescently labeled base, the emission associated with one of the bases was detected. The read out was a series of fluorescence emissions that coupled together to provide a sequence of the ssDNA.[13, 14] Many of these technologies rely on the use of fluorescent probes and imaging technologies to identify sequences.[15] The use of TIRF microscopy and fluorescent probes serves as the inspiration for the protein/peptide-sequencing platform presented in this work.

3.1.4 Marcotte Single-Molecule Protein Sequencing (Marcotte Sequencing)

TIRF microscopy provides a level of sensitivity that exceeds mass spectrometry. This increased sensitivity is due to the TIRF platform using an evanescent wave of electromagnetic energy that radiates from the interphase between the glass slip and oil-phase. Only molecules within 200 nm from the interphase will only be excited. Thus, any fluorescent impurities that may be found in the bulk solution do not interfere with the compounds to be sequenced. Unlike DNA proteins do not have the machinery available to amplify the sequence, limiting the ability to improve signal detection by amplifications like PCR. Therefore, in order to extract sequence information, biochemists have traditionally relied on Edman degradation (Figure 3.1). This chemistry removes the N-terminal amino acid residue in a step-wise fashion. Usually, reverse-phase separation and detection of elution times by UV-Vis is required to identify the residue removed. This method for sequencing proteins predated the use of mass spectrometry, and like mass spectrometry this method cannot resolve complicated mixtures of peptides.[16]

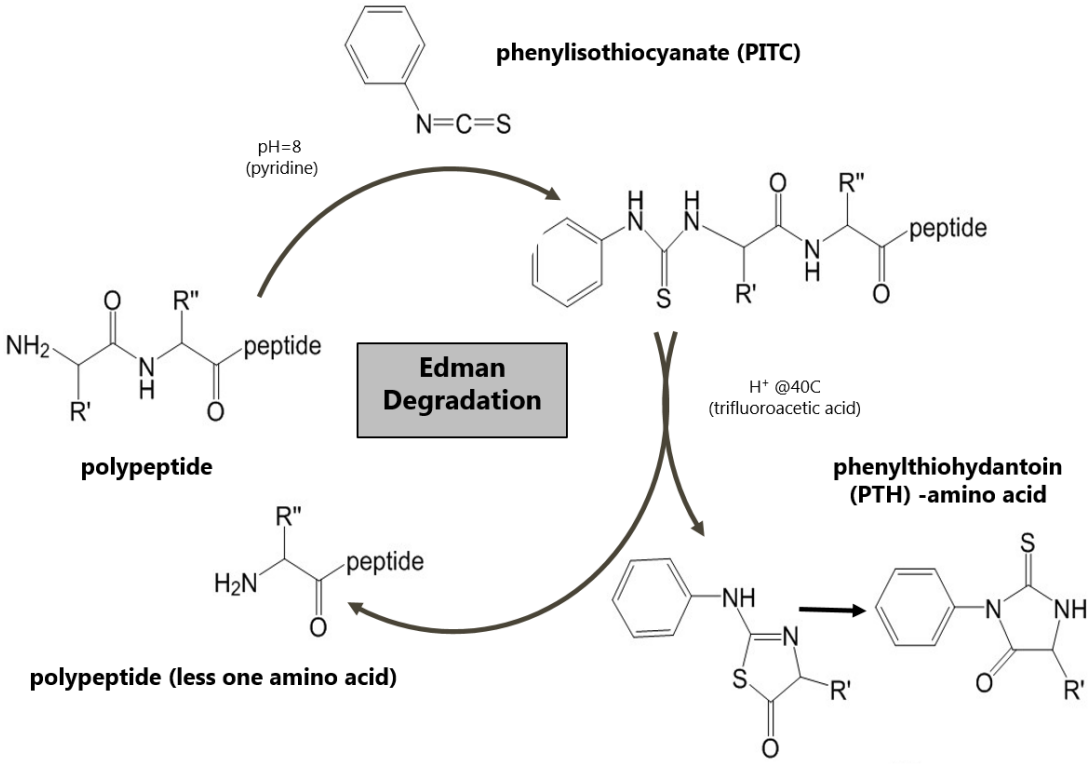


Figure 3.1: Edman Degradation Cycle. Only-N-terminal amino acid is removed. (Image credit: Angela M. Bardo (Marcotte).

Edman degradation combined with TIRF microscopy is the pivotal feature of the Marcotte sequencing platform. As described in Figure 3.2, the overall aim is to extract proteins from a sample: tissue, blood, saliva, etc. The proteins will be digested using site-specific proteases. These peptide digests will be labeled with amino acid-residue specific fluorophores from the side-chain selective functional handles described in Chapter 2. After being labeled, these peptides will be immobilized by covalent attachment on a glass slip. Edman degradation will be performed on these peptides. The loss of fluorescence

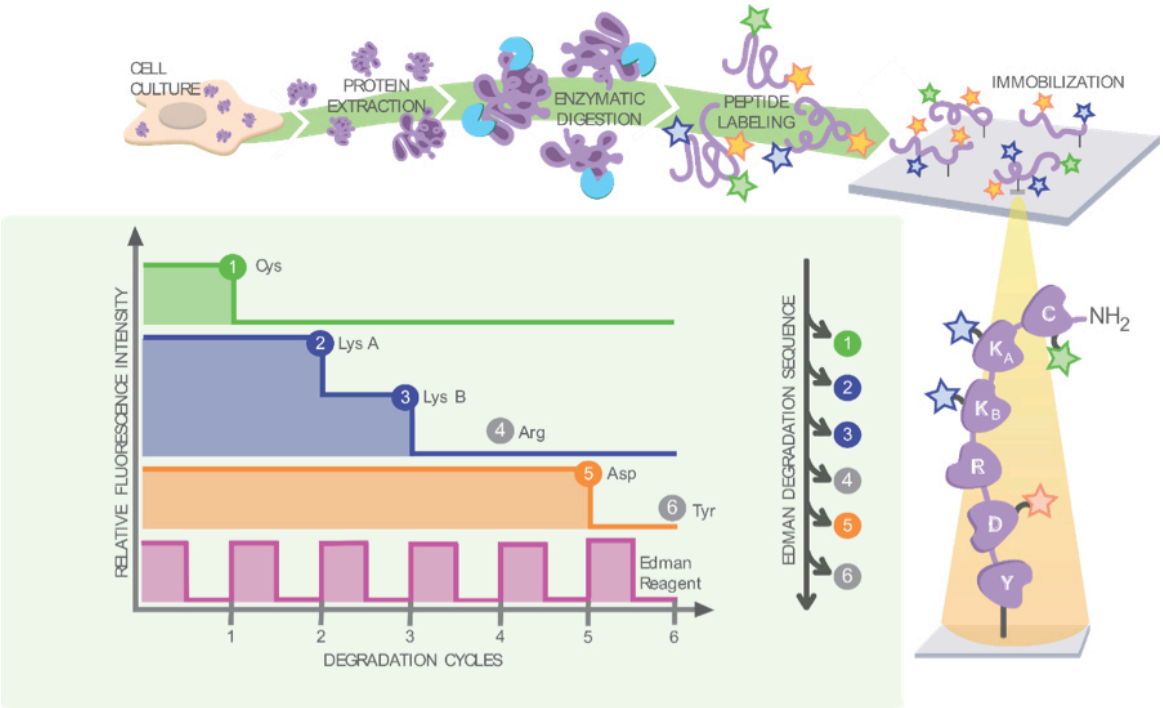


Figure 3.2: Schematic for the overall goal for the Marcotte sequencing approach. (Image credit: Angel Syrett).[9]

intensity after these degradation cycles (Edman cycles) will be monitored. For every residue labeled, a different emission wavelength will be monitored. When the intensity for a particular wavelength is observed, the identity along with the position of an amino acid can be determined, resulting in a partial sequence.

This partial sequence can be searched against protein databases to assign the protein origin of the peptide fragment. Since this is a single-molecule technique, this approach is order of magnitudes more sensitive than mass spectrometry, because it is a single-molecule approach. Furthermore, the technology described is also a massively parallel approach, increasing the sequencing throughput similar to next-generation DNA sequencing (Figure 3.3).

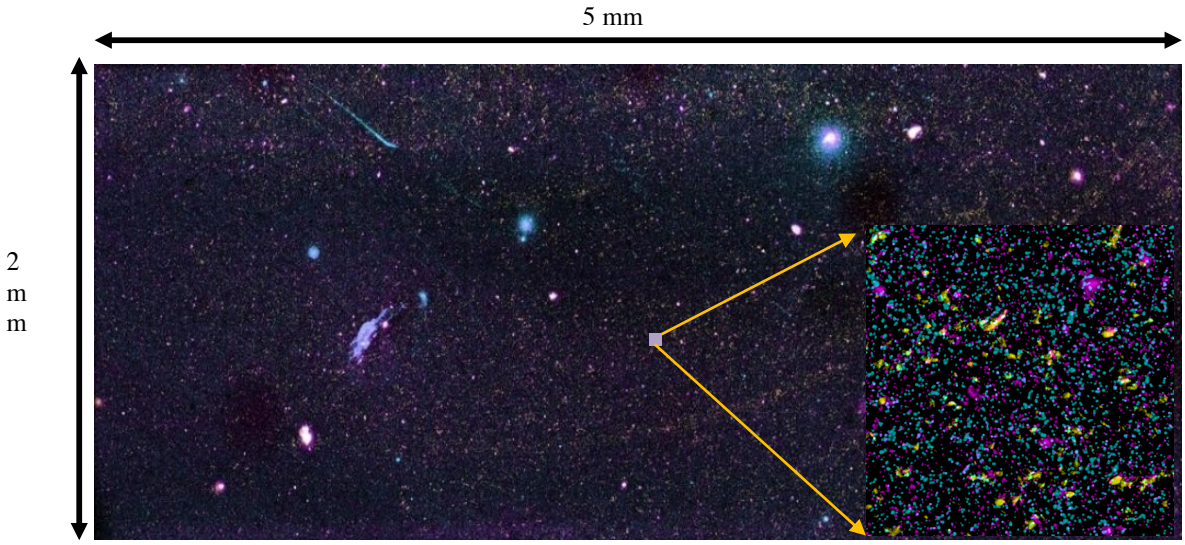


Figure 3.3: Fluorescently labeled peptides on a glass surface. Yellow spots are gold nanorods that serve as fudicial markers. Magenta spots are peptides visible in the red channel. Cyan spots are a peptide visible in the red channel. This figure is a composite of 4000 fields stitched together containing 1.7 million spots. (Image credit: Jagannath Swaminathan (Marcotte))

3.1.5 Theoretical Justification for Marcotte Sequencing

The Marcotte group recently published work explaining the justification for this approach.[9] In the most ideal case, 98% of the human proteome is possible within ten Edman cycles. This high coverage was possible despite not labeling all naturally occurring residues. Factors, such as choice of side chains to label, site-specific cleavage, and anchoring of the peptide on the glass cover slip also play an important role when identifying a protein. Even introducing errors introduced into the modeling, such as inefficient Edman degradation, blinking events, and photobleaching, reasonable coverage of the human proteome was still observed (Figure 3.4).

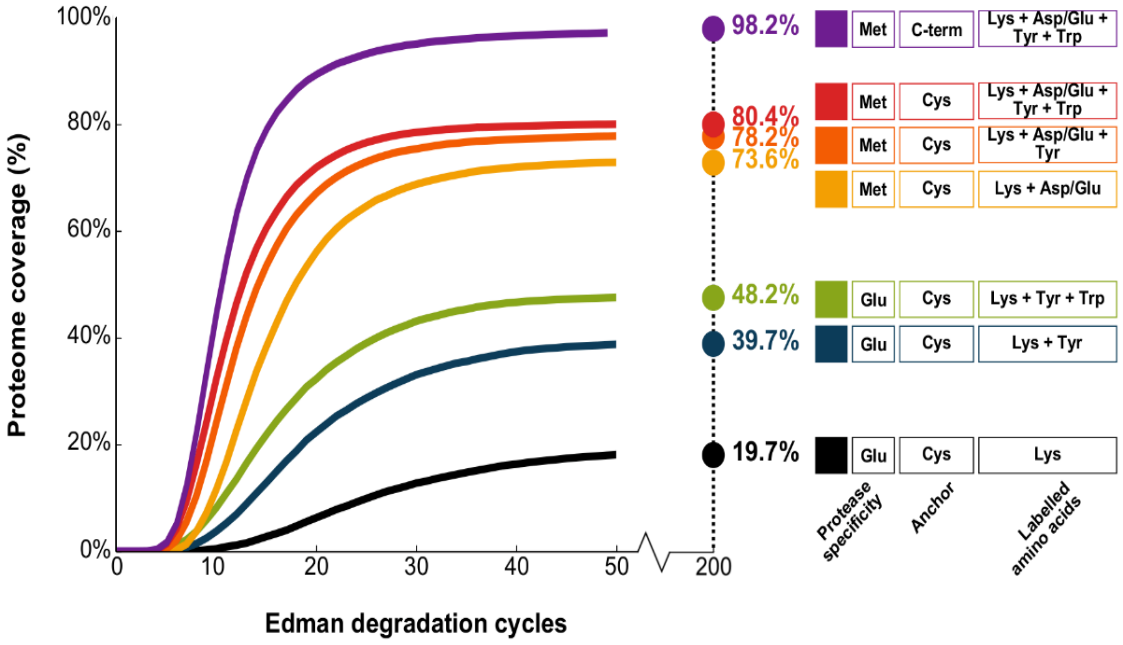


Figure 3.4: Ideal coverage of the human proteome under different cleavage conditions, peptide immobilization anchoring points, and side chains targeted. (Image credit: Alexander M. Boulgakov (Marcotte)).[9]

3.1.6 Competing Single-Molecule Protein Sequencing Approaches

Other groups are currently working on their version of single-molecule protein sequencing, most notably the Joo group at the Delft University of Technology. Theoretical work justifying their technology was recently published.[17] Their approach focuses on the translocation of fluorescently labeled proteins through a fluorescently labeled nanopore. FRET occurs as the labeled side chain exits the pore, and this event is used to identify the residue and its position along the protein. Unlike Marcotte sequencing, there is no need to digest the proteins, but the method does rely on the labeling of cysteine and lysine residues, as described in Chapter 1. In addition, the Joo approach to sequencing lacks the massively parallelize component emphasized in this work. Despite these limitations, this alternative

technology is also seen as a strong contender in the demand for new protein sequencing technologies.

3.1.7 Model Peptides Required for Sequencing

In this chapter, the design, synthesis, and sequencing of fluorescently labeled model peptides for proof-of-concept studies is described. First, peptides used for sequencing in bulk fluorescence experiments were made to evaluate if Edman degradation chemistry was possible in the presence of sterically large side chains. A screening process for dyes that survive the stringent conditions of the Emdan cycles was performed to identify robust fluorophores. Different fluorophores appended to peptides were synthesized, and tetramethyl rhodamine (TMR) (Figure 3.5a) was the dye of choice due to its stability, relative inexpensiveness, and ease to use during synthesis. After completing these bulk studies, peptides for single molecule sequencing were required. Atto 647N (Figure 3.5b) labeled peptides were preferred due to improve signal to noise and photostability relative to TMR. Natural peptides were synthesized to demonstrate that proteins in nature can be labeled, and a partial sequence can be derived from Marcotte sequencing, followed by mapping that partial sequence back to its protein origin. A synthetic digest of insulin, simulating cleavage with Glu-C, was made and labeled with Atto 647N. Also, Somatostatin-14 and a cellulomonas peptide fragment were made to demonstrate that enough sequence information was available to identify uniquely the protein origin for these simple and naturally occurring peptides. Currently, modification of recombinant insulin is being pursued as our first protein target to sequence using this platform.

Finally, in our efforts to find suitable dyes for Edman degradation, a silicone-based Rhodamine was identified as a synthetically accessible long-wavelength dye. This

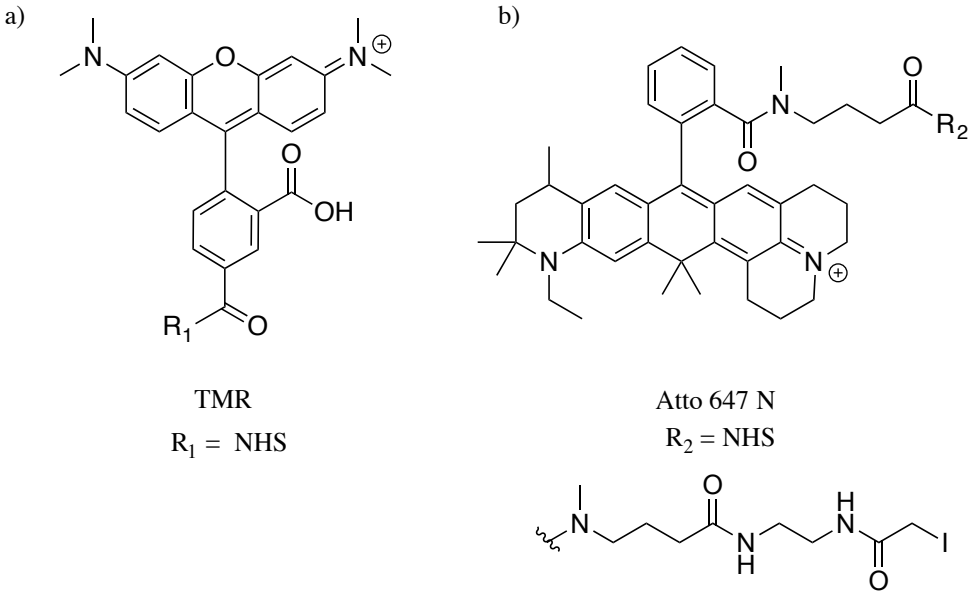


Figure 3.5: Rhodamine Dyes. a) Tetramethyl Rhodamine (TMR). b) Atto 647 N in both NHS and iodacetamide form.

dye was used to modify the N-termini of model peptides for binding studies of peptides to DNA.

3.2 Results and Discussion

3.2.1 Description of Instrumentation and Data Analysis

The microscopy, flow cell, fluidic/pump solvent delivery, automation, and image data analysis was developed by the Marcotte group (Figure 3.6).[17] Collaboration with the Anslyn group was geared towards labeling peptides with fluorescent probes and optimization of surface Edman degradation chemistries. For model studies, peptides were immobilized at the C-terminus on an amino-silane coated glass surface via amide coupling. A four-wavelength laser source was used to excite at four distinct wavelengths. Peptides attached on the surface of the glass were excited by the evanescent wave of energy emitted when light bends at the interface. Light emitted from the probes was collected by a CCD

camera and were represented as small white dots on a screen. Each white spot represented a single peptide molecule.

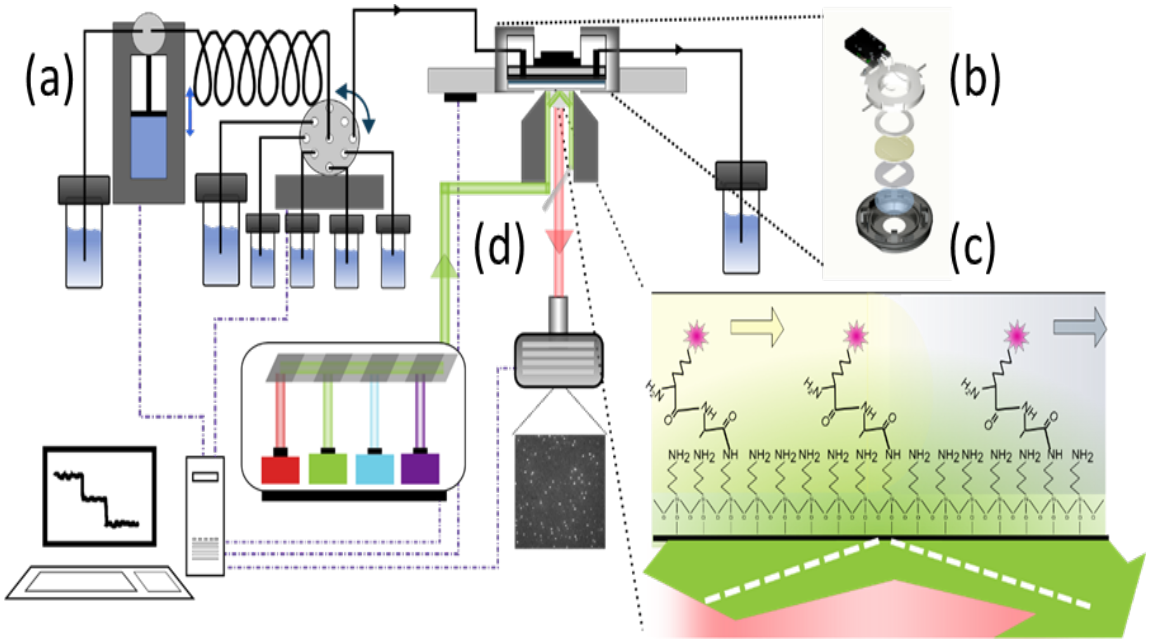


Figure 3.6: Marcotte Sequencing Platform. a) Fluidic pumps that dispense Edman degradation reagents and reactants. b) Flow cell composed of organic compound resistant materials. c) Glass slipcover functionalized with amino silanes. Peptide immobilization and Edman degradation occurs here. d) The inverted objective used to focus laser light and collect fluorescence emission captured by the camera. (Image credit: Jagannath Swaminathan)

Initially, phenyl isothiocynate (PITC) was omitted to perform three "mock" Edman cycles. These cycles were controls for fluorescence loss due to solvents and bases used for degradation. These cycles also helped to remove any fluorescent impurities and non-specifically bound peptide at the surface. After these mock runs, Edman cycles were performed as required to go through the model peptide. Images were taken after each cycle to monitor fluorescence loss. Once the experiment was completed, these images were aligned against inert reference points, such as gold nanoparticles located at the surface, to

ensure the same spot was correctly identified for each cycle. Once the images were aligned properly, the data analysis tracked the changes in fluorescence intensity for each spot.

Fluorescence intensity loss at different Edman cycles, is displayed as histograms (Figure 3.9) or heat maps (Figure 3.10). These plots counted the spots that experience significant fluorescence loss after an Edman cycle.

3.2.2 Bulk Fluorescence Studies

Before performing single-molecule experiments, a bead assay was developed to test if sequencing was possible by Edman degradation and monitoring fluorescence loss. These studies were referred to as bulk studies, since individual molecules were not imaged. Instead, the fluorescence was observed on a tentagel polystyrene bead functionalized with amines. Model peptides were immobilized on the bead via the C-terminal end using EDC coupling conditions. Peptides 3.1-3.12 (SI) were synthesized for these preliminary studies. Similar to the description provided in section 3.2.1, beads were placed under a TIRF microscope and the fluorescence was imaged after each Edman cycle. Unlike section 3.2.1, Edman chemistry was performed manually with fritted syringes. Portions of the beads were removed and imaged after each degradation.

3.2.3 N-terminal Protecting Group and Dye Selection for Bulk Fluorescence Studies

Initially, peptide 3.1 was synthesized for the bead assay. This peptide was made using a similar N-terminal protecting group from Chapter 1, known as 1-(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl (ivDde). A Lys building block was commercially available with this protecting group. However, upon the deprotection of the N-terminus, with 2% vol. hydrazine in DMF, loss of fluorescence happened. Although characterization of the degraded fluorescent peptide was not performed, attack of the 1' position on the TMR by the hydrazine was seen as a likely starting point. Attempts to

reverse this loss by acidifying the beads with TFA were not fruitful, suggesting conjugation of peptide may have been lost due to the hydrazine.

As an alternative, peptide 3.2 was synthesized with a boc protecting group with Rh. Rhodamine B was used to decrease the cost of making fluorescently labeled peptides; Rh B is available in larger quantities at a cheaper price. The dye was stable to TFA deprotection required for boc removal. Keeping concentrations consistent between Peptide 3.1 and 3.2 when immobilizing peptide on the bead, peptide 3.2 consistently led to a less bright bead.

To improve brightness of peptides, peptides 3.3 and 3.4 were synthesized. Rh 101 dyes are known for their higher quantum yields relative to TMR and other rhodamine-based dyes.[18] Structural constraints of this fluorophore are due to the julolidine moiety present, which contributes to this dyes improved quantum yield. Again, despite the literature precedent for improved brightness, at similar concentration for bead immobilization, this peptide was not bright enough. An argument about the steric size of the peptides was hypothesized. Since these peptides were relatively short, and the steric size increased with Rh B and Rh 101, coupling efficiency contributed to decrease in the brightness of beads.

Since loss of fluorescence was caused by hydrazine, TMR was tested against TFA and piperidine. The dyes appeared to be stable after treatment with these solvents and considered suitable for removing boc or Fmoc protecting groups post immobilizing model peptides on beads. Therefore, Fmoc-K*A, Fmoc-GK*A and Fmoc-K*AK*A (Peptides 3.5-3.7 where * = TMR) were used for the bead assay. Fmoc was chosen as the protecting group, because the N-terminal amino acid can be left unprotected during solid-phase peptide synthesis. Placing the boc group would add an additional synthetic and purification step.

3.2.4 Bead Assay Results

As shown in Figure 3.8, peptides 3.5 -3.7 (SI; Fmoc-K*A, Fmoc-GK*A, and Fmoc-K*AK*A, respectively, where * = TMR) were used to demonstrate the feasibility of sequencing by monitoring fluorescence loss after Edman degradation. These three peptides were synthesized to demonstrate degradation at the N-terminal position (first position), degradation at the second position, and degradation at the second and fourth position. The fluorescence data suggested peptides could be deprotected with piperidine with no adverse affect on the fluorophores. To control for fluorescence loss due to Edman degradation solvents, a mock cycle was performed. Fluorescence remained the same after this cycle. Fmoc-K*A after the first Edman cycle demonstrated fluorescence loss, and Fmoc-GK*A showed loss after the second Edman cycle. Fmoc-K*AK*A, losses were observed after the second and third Edman cycle.

3.2.5 Solution-Phase Confirmation of Edman Degradation with Fluorescent Side Chains

After successfully completing these bead assays, solution-phase Edman degradation was performed as further proof that Edman occurred with sterically large, fluorescent side chains. Peptide 3.8 (SI) was synthesized and isolated as the uncyclized, thiourea intermediate before degradation, proving N-terminal modification was possible. The isolated peptide was subjected to a solution of TFA (90% vol. in H_2O). The TFA was removed by rotary evaporation and remaining solution was submitted for high-resolution mass spectrometry (HRMS). The phenylthiohydantoin (PTH) derivative (peptide 3.8) was observed with the expected fluorescent side chain. These results served as further evidence for deriving sequence information by Edman degradation and monitoring fluorescence loss.

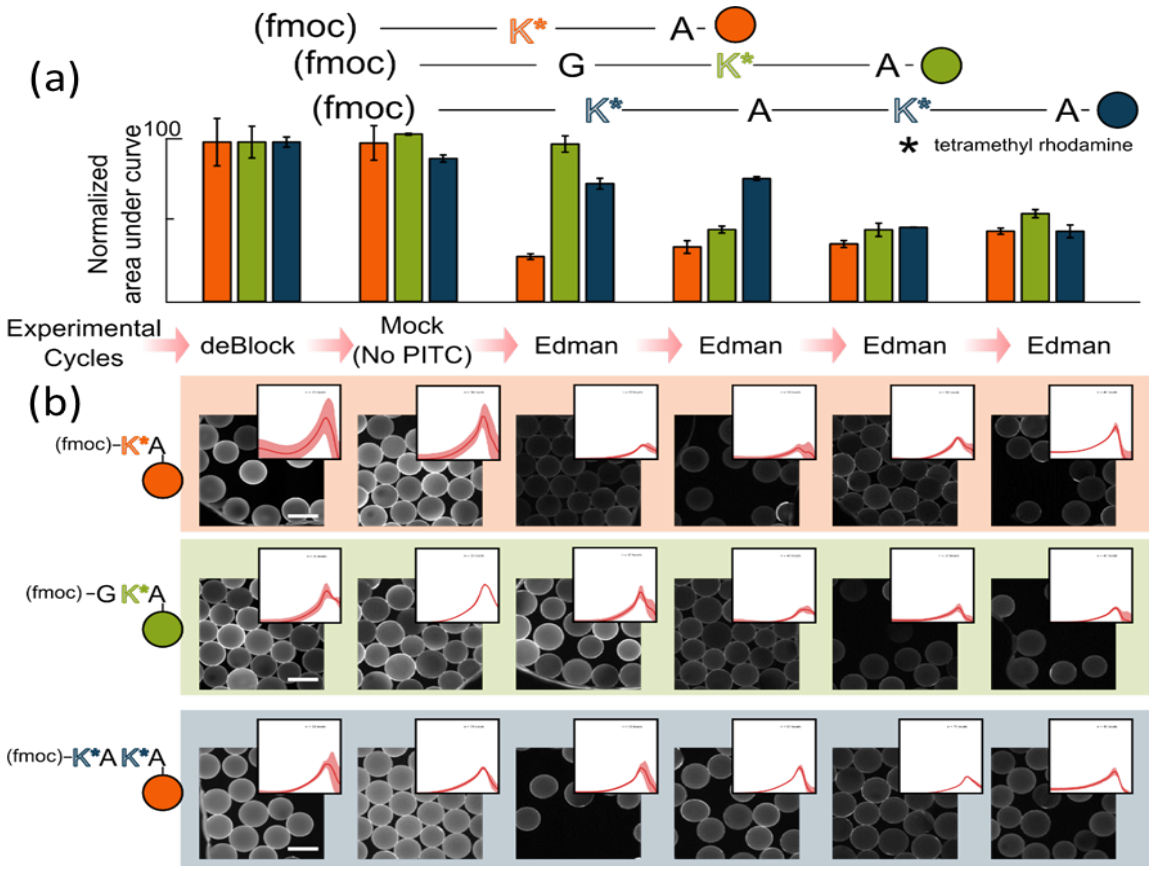


Figure 3.7: Bead Peptide Sequencing Study. a) Histogram representation of cycles performed on peptides. b) Images taken after each cycle and respective radial fluorescence intensity profile across beads. (Image credit: Jagannath Swaminathan)

3.2.6 Miscellaneous Fluorescent Peptides

Peptides 3.10-3.12 were synthesized taking into consideration what sequences can be used in future bead assays. For example, peptide 3.10 was synthesized to see if immobilization on amine surfaces was possible if cleavage of proteins were to be done with cyanogen bromide, which is site-specific for methionine.[19] Thus, all peptides were terminated with a methionine. Exploring the stability of this residue at the C-terminus when

performing Edman degradation should be done. Stability studies are important since cleavage at this residue provided the best coverage of the human proteome.

Peptide 3.11 was synthesized implementing only solid-phase chemistries. The Anslyn group synthesized a boc-protected lysine building block with a RhB. The peptide had a free cysteine as an alternate cite for immobilization. However, due the dubious quality in the brightness of RhB, further exploration of this peptide was not pursued.

Finally, peptide 3.12 was designed as a preliminary attempt to begin sequencing peptides with unnatural side chain residues. 4-ethynylbenzohydrazide (compound provided by the Anslyn group) was clicked onto an azido lysine peptide in the solid-phase. Upon cleavage and isolation, the hydrazide peptide was labeled similarly to lysine residues. Efforts to sequence this unnatural peptide are ongoing.

3.2.7 Smaller Scale Reactions

As experiments transitioned towards single-molecule sequencing, the need for Atto 647 N peptides became more important. Relative to TMR, Atto 647 N demonstrated better photostability. Also, since this dye absorbed and emitted at longer wavelengths, fewer fluorescent impurities were observed, improving the signal to noise. Atto 647 N and its derivatives were synthetically challenging to make and expensive. Therefore, minimizing the starting material required to make model peptides was an important goal. During HPLC purification the peptide was readily detected at 0.1mg/mL concentrations.

3.2.8 Atto 647 N Model Peptides Single-Molecule sequencing: Single Dye

Peptides 3.13 and 3.14 (boc-GK**AGAG and Ac-GK**AGAG, respectively, where ** = Atto 647N) were made to study sequencing of a peptide with a single dye. The peptide was labeled at the second position to ensure degradation was the cause for desired loss of fluorescence. In addition, glycine and alanine residues were selected to minimize any potential issues that arose from any steric hindrance from bulkier side chains. An

acetylated version of the GK**AGAG was synthesized to act as a control. The N-terminal amino acid was permanently protected degradation. This peptide was subjected to the same experimental run as the free amine peptide on a separate glass cover slip to take into account losses in fluorescence intensity due to boc-deprotection, non-specific solvent-induced degradation, non-specific binding to glass surface, poor glass surface quality, photobleaching, and photoblinking events. Like the bead assays, both the sequence-capable peptide and the control peptide underwent a series of mock cycles as further controls to ensure true sequencing events were identified.

Shown in Figure 3.9 is a histogram of the peptides tracked over the sequencing run. After three mock cycles, only a small percentage of the peptides tracked had complete loss of fluorescence. Losses for Ac-GK**AGAG and GK**AGAG were the same. Once the first Edman cycle was performed, no appreciable loss was observed as expected. However, after the second Edman cycle, GK**AGAG had significant fluorescence loss. In contrast, the acytelated peptide had a loss of fluorescence similar to previous cycles. Two additional Edman cycles were performed to see if more significant losses occurred. GK**AGAG did not have significant losses in fluorescence intensity after the second degradation.

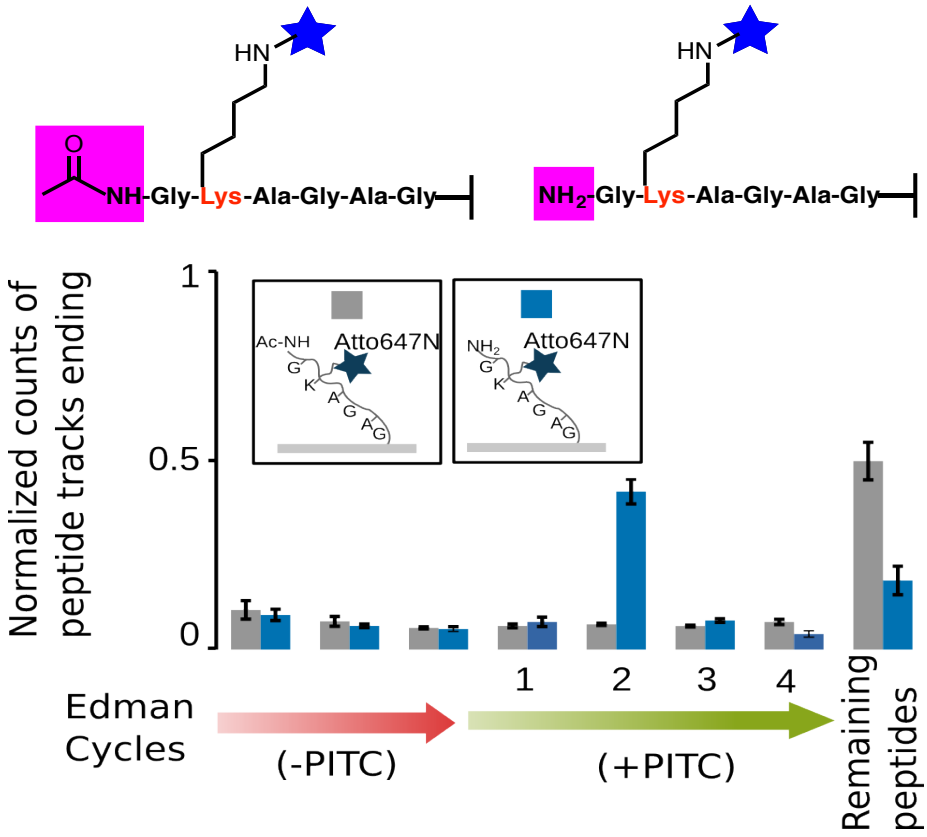


Figure 3.8: Single-molecule sequencing results for peptides 3.13 and 3.14. (Image credit: Jagannath Swaminathan)

3.2.9 Atto 647 N Model Peptides Single-Molecule sequencing: Two Dyes

Peptides with two dyes along the sequence were synthesized to study if two probes of the same fluorescence, incorporated into same peptide, could be resolved. Peptides 3.15 and 3.16 (Boc-GC**AGC**AGAG and Ac- GC**AGC**AGAG, respectively, where ** = Atto 647 N) were synthesized to monitor if the thioether linkage was able to survive Edman degradation, and to facilitate the synthesis of model peptides using the chemistries explored in Chapter 1.

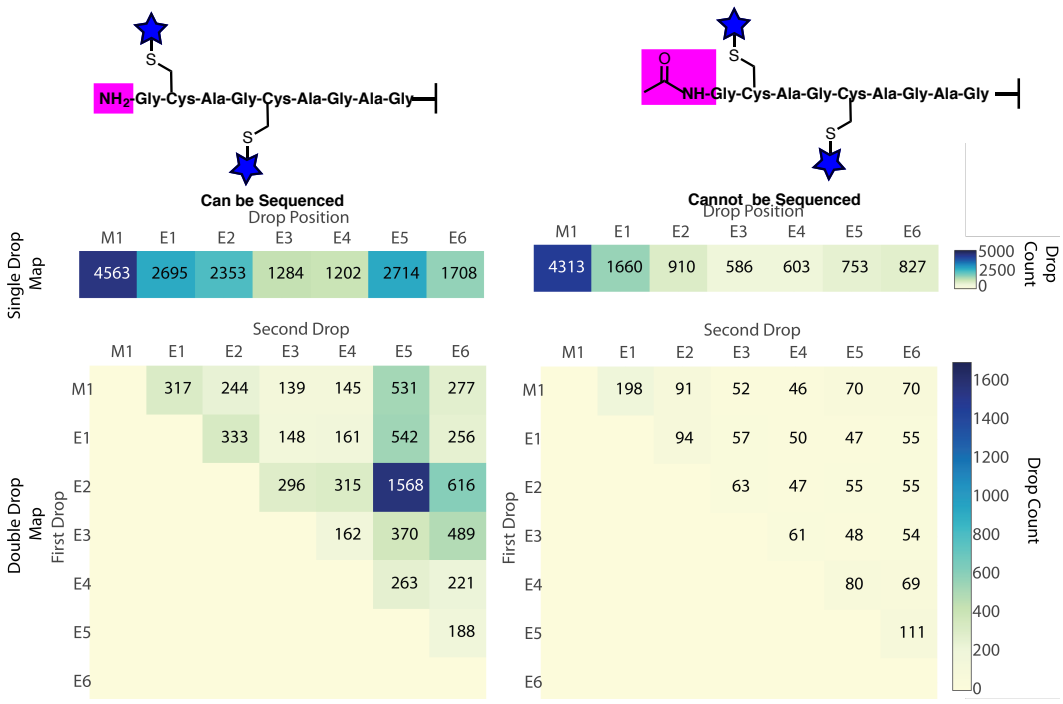


Figure 3.9: Single-molecule sequencing results for peptides 3.15 and 3.16. (Image credit: Angela Bardo).

Figure 3.9 tallies the spots monitored over one mock and six Edman cycles. The number located at the intersection of E2 (first drop) and E5 (second drop) represents the number of spots with fluorescence losses that occurred at degradation cycles 2 and 5. In other words, after the fifth Edman cycle, of the total number of spots tracked throughout the experiment, 1568 were no longer observed. The same number of spots was observed after the second cycle, but at a lower intensity. As expected, the position on the heat map with the greatest intensity is at E2 and E5, showing that true sequencing events were observed. The results also demonstrated that losses at other junctures could be identified. The sequence heat map for Ac-GC**AGC**AGAG also showed losses in fluorescence intensity. However, the number of fluorescence loss compared to GC**AGC**AGAG are still fewer, suggesting inefficiencies in Edman degradation also misidentified sequencing

events. Despite these issues, the Marcotte platform was able to identify the sequence of a doubly labeled model peptide.

3.2.10 Future Sequencing Studies: Spacing Limitation for Fluorescently Labeled Side Chains

Peptides 3.17-3.31 (SI) were synthesized as future models to explore the ability of this sequencing platform. For example, peptides 3.17-3.18 will be used for determining at what distance fluorophores on doubly labeled peptides can be distinguished. These include a peptide with two adjacent fluorescently labeled peptides (peptide 3.17) and one with two residues spaced between labeled side chains (peptide 3.18).

3.2.11 Future Sequencing Studies: Naturally Occurring Peptides

Peptides 3.20-3.21 were synthesized to simulate Somatostatin-14 (peptide 3.20) and a peptide found in the bacteria *cellulomonas*. Somatostatin-14 contains two cysteine residues and two lysine residues. However, cysteine was not labeled with a fluorescent probe, because models showed that it can still be identified uniquely against the human proteome with only Lysine residues labeled.

3.2.12 Future Sequencing Studies: Synthetic Insulin Digests

Peptides 3.22-3.29 were synthesized to simulate insulin digested with Glu-C, which cleaves specifically after carboxylate residues, and resulted in 6 peptide fragments. Four can be labeled with iodoacetamide dyes (Figure 3.11). Along with their acetylated counterparts, these peptides will be used to determine if insulin can be uniquely identified using the Marcotte sequencing platform.

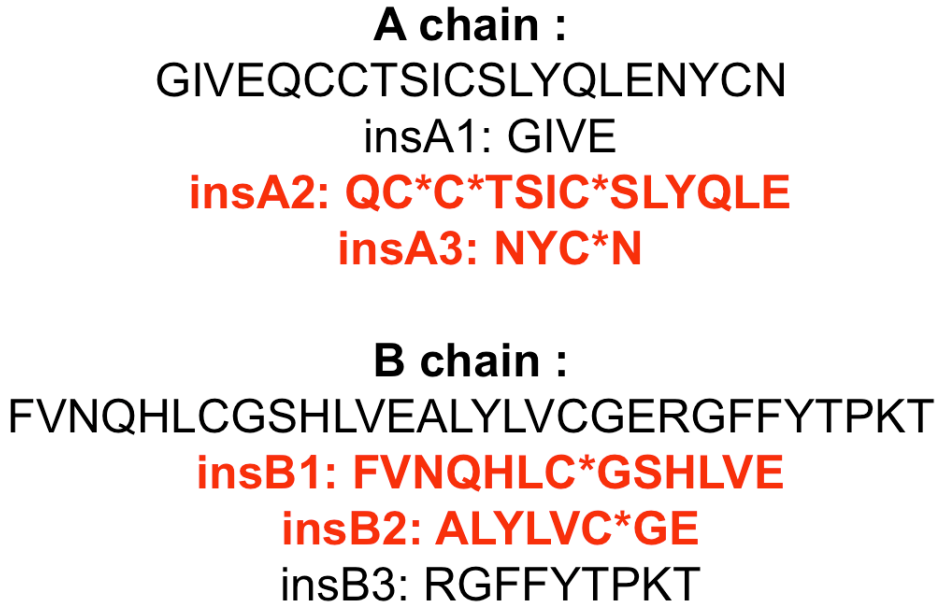


Figure 3.10: Insulin (ins) fragments post Glu-C digestion. From chains A and B, four peptides can be labeled at cysteine. Each peptide gives a unique partial sequence which can be uniquely identified to insulin when searched against the human proteome.[20]

3.2.13 Labeling of Cysteine Residues in Recombinant Human Insulin

The ultimate application for the Marcotte sequencing technology will be to identify proteins and peptides from biological samples. Thus, our focus is aimed towards recombinant insulin. This serves as a next step in labeling studies inching closer to our ability to label protein and peptide mixtures with fluorophores. Insulin was selected because of its commercial availability and cost.

In our efforts to digest insulin, reduction of the disulfide bonds joining chains A and B was performed by addition of TCEP. Insulin was difficult to dissolve at the concentrations used, but it was possible with 14 N NH_4OH. The stoichiometry of the reducing agent was also controlled during reduction, three equivalents for each disulfide bond present. Excess TCEP would react with the iodoacetamide dye introduced. TMR was

used instead of Atto 647 N because of the cheaper cost. In future studies, chemistries to optimize the labeling of insulin with Atto 647 N will be pursued.

HPLC purification was first attempted to isolate the peptides from the hydrolyzed dye. The run proved unsuccessful since colored fractions did not contain desired products. The samples were difficult to isolate, because the peptides were quite large and hydrophobic. These two factors contributed to the difficulty in solubilizing modified compounds in aqueous mixtures. Sodium dodecyl sulfate (SDS) was used to promote solubility prior to purification. Product also was filtered from the solution when passed through a 0.2 micron syringe filter. Washes with H_2O and MeCN did not entirely remove chains A and B. Peptides were washed off using DCM. No further purification was attempted. The filtering unintentionally served as crude and effective means to remove most of the hydrolyzed dye impurity (Figure 3.12 and 3.13).

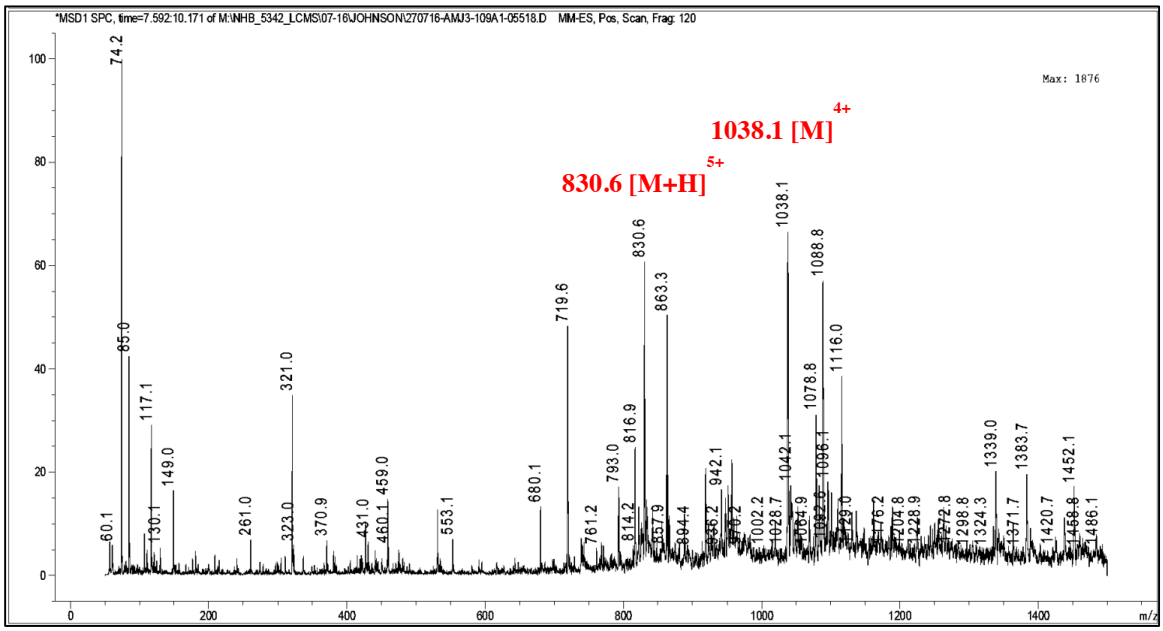


Figure 3.11: Insulin Chain A labeled with TMR isolated by syringe filter.

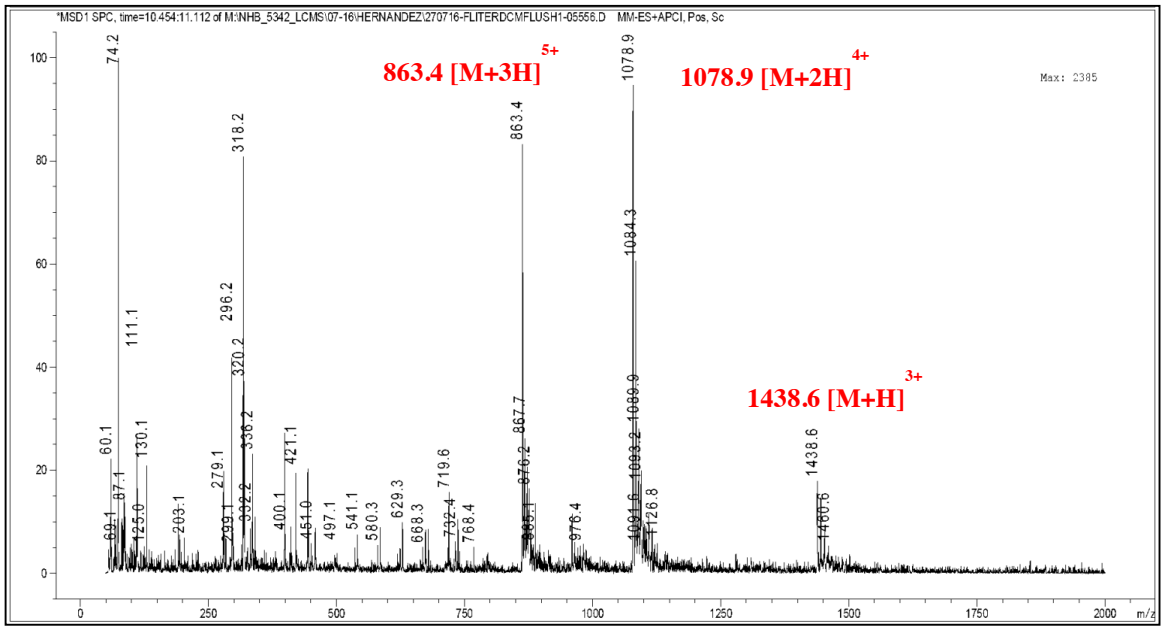


Figure 3.12: Insulin Chain B labeled with TMR isolated by syringe filter.

3.2.14 Digestion of TMR Labeled Insulin Fragments

After modification to the procedure, studies to see if peptides could be digested with bulky fluorescent probes were pursued. The labeled A and B chains were dissolved in a 1:1 MeCN:H_2O aqueous mixture. Ammonium bicarbonate (100mM) was the buffer used to selectively cleave the glutamates residues.[21] Glu-C was introduced and incubated at RT at 37 °C overnight, and submitted for LCMS. Fragments from chain B were seen more readily, because of protonatable sites such as histidine and smaller masses that increased the ionization efficiencies compared to the more hydrophobic, heavier insulin A fragments. Figures 3.14 and 3.15 summarize the masses observed.

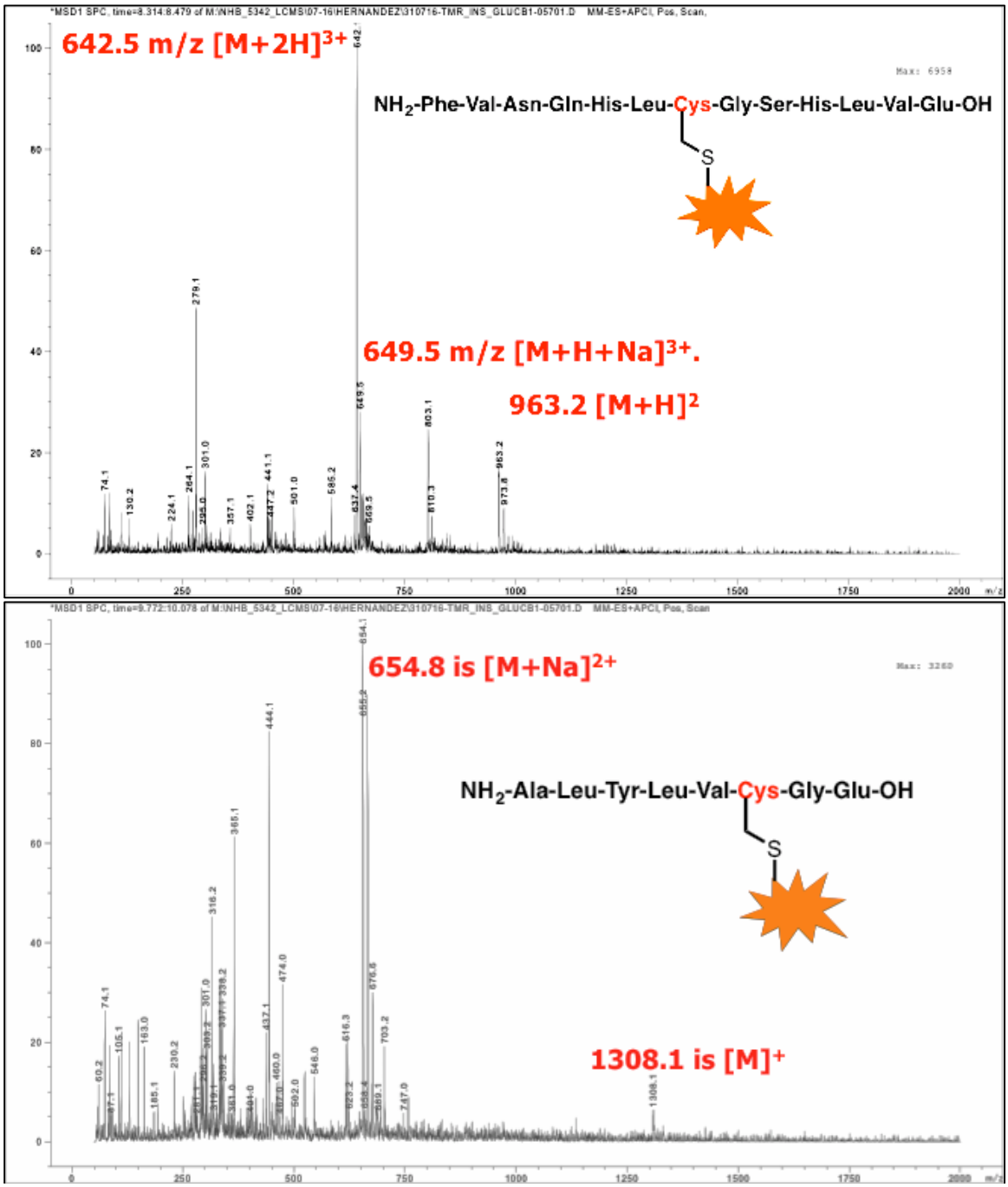


Figure 3.13: LRMS Ins. Fragments. (top) TMR labeled insB1. (bottom) TMR labeled insB2.

Current aims will be to transition this labeling chemistry to Atto 647 N and digestion with Glu-C followed by sequencing of recombinant insulin. Obstacles include the hydrolysis of the dye, which happens more readily than TMR. Attempts to decrease the pH and increasing the amount of organic solvent for solubilizing are expected to mitigate hydrolysis. As with the TMR, labeling protein digests with Atto 647 N increased their hydrophobicity. Increasingly hydrophobic peptides create difficulty when trying to dissolve both Glu-C and peptides. Proteases, like Glu-C, at most tolerate 20% v MeCN. Solubility should be optimized when labeling before continuing to label mixtures of proteins.

3.2.15 Silicone-based Rhodamine for N-Terminal Modification (Si-Rh)

During the screening of dyes, the Anslyn group provided a silicone-based Rhodamine (Si-Rh) dye as a cheaper alternative to Atto 647 N (Figure 13.15). Unfortunately, this Rhodamine dye was not stable to Edman degradation conditions. The Si-C bond was presumed to be too labile to the conditions required for sequencing. Nevertheless, applications that require imaging in aqueous environments have utilized

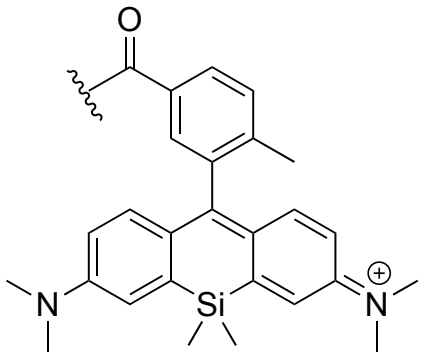


Figure 3.14: Silicone-based rhodamine (Si-Rh).

the dye.[22] A series of peptides modified at the N-terminus incorporating Si-Rh were synthesized for binding studies to DNA. A general template was followed when synthesizing peptides: Si-Rh–DYKDDDDKXAXAXA, where X = one of the 20 naturally occurring amino acids. Changing the residue will help determine what sequences of DNA bind preferentially to these amino acids. DYKDDDDK, known as Flag Tag, was included as an anchoring point to isolate these peptides via antibodies. These experiments are ongoing and no data is available.[23]

Conclusions 3.3

Single-molecule peptide sequencing studies were conducted by monitoring fluorescence loss of labeled peptides undergoing Edman degradation. Model peptides labeled with Rhodamine dyes were first synthesized to confirm that sequence information could be obtained by bulk fluorescence studies on beads. Atto 647 N peptides were subsequently synthesized for single-molecule proof-of-concept studies. Studies sequencing these model peptides are ongoing. Efforts are also directed to label insulin as the first protein target to be modified, digested, and immobilized on the Marcotte platform. Future directions will include the incorporation of two colors to differentiate between amino acid residues. From these studies, Si-Rh and was synthesized and incorporated to N-termini in peptides that will be used for DNA binding studies.

3.4. Experimental

3.4.1 General Materials

For automated, Fmoc amino solid-phase peptide synthesis, Ala, Ile, Leu, Met, Phe, Pro, Val, Pbf (Arg), OtBu (Asp, Glu), Boc (Lys, Trp), tBu (Ser, Thr, Tyr), Trt (Asn, Cys, His), and ivDde-Lys(boc)-OH were purchased from Novabiochem (USA), AAPPTec (USA), P3 Biosystems (USA), Chem-Impex (USA). (*S*)-*N*-(9-(2-((2-(3-((*tert*-butoxycarbonyl)amino)-3-carboxy-*N*

methylpropanamido)ethyl)(methyl)carbamoyl)phenyl)-6-(diethylamino)-3*H*-xanthen-3-ylidene)-*N*-ethylethanaminium was provided by James Logan Bachman from the Anslyn group. Fmoc-Ala-Wang resin (0.62 mmol g^{-1}), Fmoc-Met-Wang resin (0.30 mmol g^{-1}), Fmoc-Asp(OtBu)-Wang resin (0.69 mmole g^{-1}), Fmoc-Gln-Wang resin (0.47 mmol g^{-1}), Fmoc-Asn-Wang resin (0.57 mmol g^{-1}), and Fmoc-Gly-Wang resin (0.62 mmol g^{-1}) were purchased from Novabiochem (USA) and P3 Biosystems. Other chemicals used for automated, solid-phase peptide synthesis were purchased from Fisher Scientific and Sigma-Aldrich. N,N′-Diisopropylcarbodiimide (DIC) and Oxyma and (ethyl cyano(hydroxyimino)acetate) was purchased from Chem-Impex. N-hydroxy-succinimide tetramehtyl rhodamine (NHS-TMR), tetrmethylrhodamine-5-iodoacetamide dihydroiodide (5-TMRIA) was purchased from Pierce. Rhodamine B (RhB) and Rhodamine 101 (Rh101) were purchased from Sigma-Aldrich. *N*-(7-(dimethylamino)-10-(5-(((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)-2-methylphenyl)-5,5-dimethyldibenzo[*b*,*e*]silin-3(5*H*)-ylidene)-*N*-methylmethanaminium (Si-Rh) was provided by Anslyn group member James Logan Bachman. N-hydroxysuccinimide boc was purchased from ACROS organics. Boc anhydride was purchased from Fisher. Recombinant insulin, Tris-(2-Carboxyethyl)phosphine, Hydrochloride (TCEP) was purchased from Sigma-Aldrich.

3.4.2 General Instrumentation

A Prelude peptide synthesizer (Protein Technologies, Inc.) was used for automated-solid phase synthesis. For longer sequences, a Liberty Blue microwave peptide synthesizer was used for solid-phase peptide synthesis. Preparative HPLC purification of peptides was performed using an Agilent Zorbax SB-C18 Prep HT column 21.2 250 mm; 10 mL min1 , 5–95% MeOH (0.1% FA) in 90 min. Analytical HPLC characterization of peptides was performed using an Agilent Zorbax column 4.6 250 mm; 1 mL min1 , 5–95% MeCN (0.1% TFA) in 35 min (RT). A Gemini C18 3.5 micron 2.1 50 mm was used for online separation;

0.7 mL min1 , 5–95% MeCN (0.1% formic acid) in 12 min (RT) for LCMS characterization. An Agilent Technologies 6530 Accurate Mass QTofLC/MS mass spectra of purified peptides.

3.4.3 General Solid-Phase Model Peptide Synthesis

Peptides were synthesized using their respective resin with sequential coupling of N_{α}-Fmoc-amino acid (0.1 M, 1.5 ml) in DMF in the presence of N,N,N,N-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU, 0.15 M, 1.0 ml) and DIPEA (0.2 M, 0.5 ml) with a reaction time of 30 minutes at room temperature. A total of three repetitions were performed for each amino acid building block. DMF (3 ml, 3 min, 3x) and DCM (3 ml, 3 min, 3x) washes were done before each repetition. After incorporation of the third amino acid, a 0.8 M LiCl wash step was performed after swelling with DCM (3 ml, 3 min, 3x). Standard settings for Liberty Blue microwave peptide synthesizer were used to make longer peptides. Post synthesis, resins were washed with glacial AcOH (5 ml, 3x), DCM (5 ml, 3x), and MeOH (5 ml, 3x). The resins were placed under vacuum overnight. If cysteine was present, the peptide was cleaved using trifluoroacetic acid (TFA), triisopropylsilane (TIS), 1,2-ethanedithiol (EDT), and nanopure water (94 : 1.0 : 2.5 : 2.5), and precipitated with diethyl ether at 0 °C. No HPLC purification of the crude peptide was necessary. If cysteine was not present, TFA, TIS, and nanopure water were used (95 : 2.5 : 2.5) to cleave peptides.

Peptides labeled with hydrzinobenzoyl resins were first synthesized on the solid support as described previously. 100 mM (1.5 mL) dye of choice, 1.2 M of DIPEA (0.5 mL), and 300 mM (1.0 mL) in DMF were introduced to resin. Afterwards, resin was washed and placed under vacuum. Copper acetate (0.33 mmole) was dissolved in 3 ml of 9/8.3/1.6 MeCN/Pyr/1-amino-3-butyne (v/v/v). Subsequently, the solution was introduced to swollen resin. The resin was incubated for 4 h at RT, followed by filtration to collect the

solution, and the MeCN and pyridine were removed by rotary evaporation. Crude product was purified via preparative HPLC. Organic solvent was removed via rotary evaporation, and aqueous remnants were frozen and lyophilized overnight at -70 °C.

3.4.4 Labeling of N-terminal ivDde/Fmoc-protected peptides with Cysteine or Lysine Residues

Peptides were dissolved in DMF 50-100 μL. To this solution, 50-100 μL of DIPEA was added. Iodoacetmaide (IAc) (for Cysteine) or N-hydroxysuccinimide (NHS) activated dye (for Lysine) was added to the solution and incubated at RT overnight. Compound was purified by preparative HPLC. Fractions were collected and organic solvent was removed by rotary evaporation. Aqueous remnant were collected and frozen. Sample was lyophilized overnight at -78 °C.

3.4.5 Synthesis N-terminal boc-protected peptides with Cysteine or Lysine Residues.

Peptides were first modified with NHS-dyes as described in 4.4.4, followed by deprotection with introduction of piperidine (20% the volume of DMF present in reaction). Compound was purified and isolated by preparative HPLC and lyophilization. Purified peptide was dissolved in DMF 50-100 μL, followed by 50-100 μL of DIPEA. NHS-boc was introduced in 10 eqs. The solution was incubated overnight at RT overnight. Peptide was purified via preparative HPLC and lyophilization.

3.4.6 Solid-phase N-terminal Labeling of Peptides with NHS-Si-Rh

N-terminal protecting group for the last amino acid incorporated into the sequence was manually removed with piperidine (20% vol. in DMF, 3 mL, 3 min, 3 x). 57.5 mg of dye were dissolved in 1000 μL of DMF to make a 100 mM stock dye solution in DMF. 100 μL of a 1.2 M DIPEA solution and 300 μL of stock dye solution were mixed with 20 mg of resin overnight (16 hours) in the dark. Excess labeling solution was washed from the

resin with DMF until the filtrate ran clear. Resin was treated similarly as described in 4.4.3 post solid-phase synthesis. Peptide was isolated via preparative HPLC and lyophilization.

3.4.7 Acetylation of Peptides

Peptides were generally acetylated at the N-terminus manually by introducing 2 mL of DCM, 1 mL of 1.2 M N-methyl morphiline, and 1 mL of boc anhydride to resin with a peptide with a deported N-terminus. Mixture was incubated at RT for 2 hrs, and the solution was removed. The resin was subjected to the same reaction conditions a second time. The resin was treated post-synthesis as described in 4.4.3. Alternatively, boc anhydride could be introduced to the peptide synthesizer and treated as an additional building block in the sequence.

3.4.8 Labeling of Human Recombinant Insulin

4 mg of recombinant insulin was dissolved in 500 μL of 14 N NH_4OH, and 3 eqs. of TCEP were introduced into the solution. The solution was incubated at RT for 15 min. (5-TMRIA) was introduced at 6 eqs and incubated at RT for 3 hrs. Crude isolation of peptide was performed by passing solution through a Fisherbrand 13 mm syringe filter (0.22 μm, PVDF). 1000 μL of 1:1 H_2O/MeCN was passed through filter, followed by 1000 μL of DCM. Organic solvent was removed by rotary evaporation.

3.4.9 Glu-C Digestion TMR Labeled Chain A and Chain B

Crude sample isolated from section 4.4.8 was dissolved in 500 μL of 1:1 (H_2O/MeCN). 2000 μL of an ammonium bicarbonate solution was introduced, followed by 90 μg of Glu-C. Sample was not isolated but submitted for LCMS to identify peptide fragments.

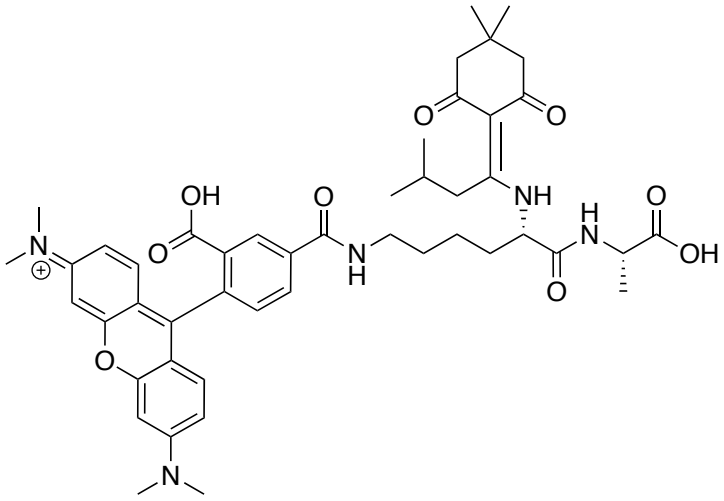
3.5 Acknowledgements

We gratefully acknowledge our funding sources: the National Institute of Health (5DP1GM106408), the Welch Foundation (F1515 to E. M. M.), the Defense Advanced Research Projects Agency (DARPA, N66001-14-2-4051), and the Welch Regents Chair to EVA (F-0046).

We also gratefully acknowledge James Logan Bachman for providing Rhodamine dyes and fluorescently labeled amino acid building blocks. Special mention goes to Cedric Ginestra, an Anslyn undergraduate researcher, who greatly contributed to the design and synthesis of Si-Rh modified peptides.

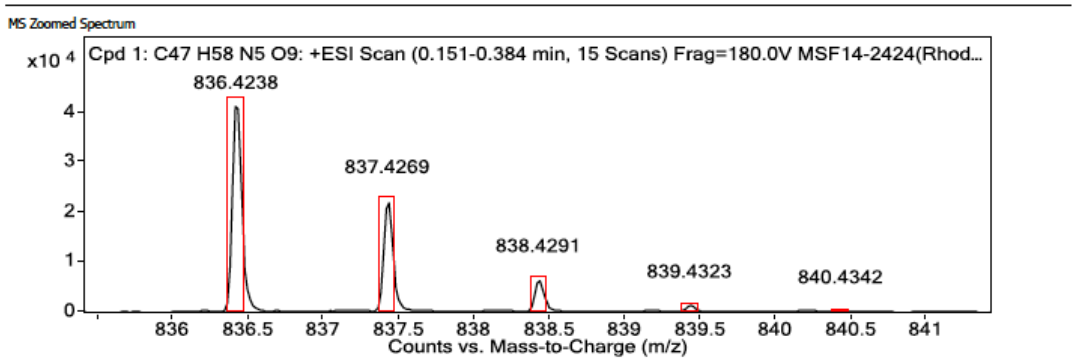
From the Marcotte group, we acknowledge Jagannath Swaminathan, Alex Boulgakov, and Angela Bardo for their figures, which made writing this chapter more clearly.

3.6 Experimental Characterization



m/z: 836.42 (100.0%), 837.43 (50.8%), 838.43 (12.6%), 838.43 (1.8%), 837.42 (1.8%), 839.43 (1.2%)

ivDde-lys(TMR)-ala



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
607.15810			74689.15			
836.42380	836.42290	1	42859.08	C47H58N5O9	M+	-1.09
837.42690	837.42610	1	22271.77	C47H58N5O9	M+	-0.92
838.42910	838.42900	1	6478	C47H58N5O9	M+	-0.1
839.43230	839.43180	1	1331.11	C47H58N5O9	M+	-0.62
840.43420	840.43460	1	200.05	C47H58N5O9	M+	0.46
841.42570	841.43720	1	40.9	C47H58N5O9	M+	13.66

Figure 3.15: HRMS data for peptide 3.1. Starting amount: 0.012 mmole. Yield: 33%.

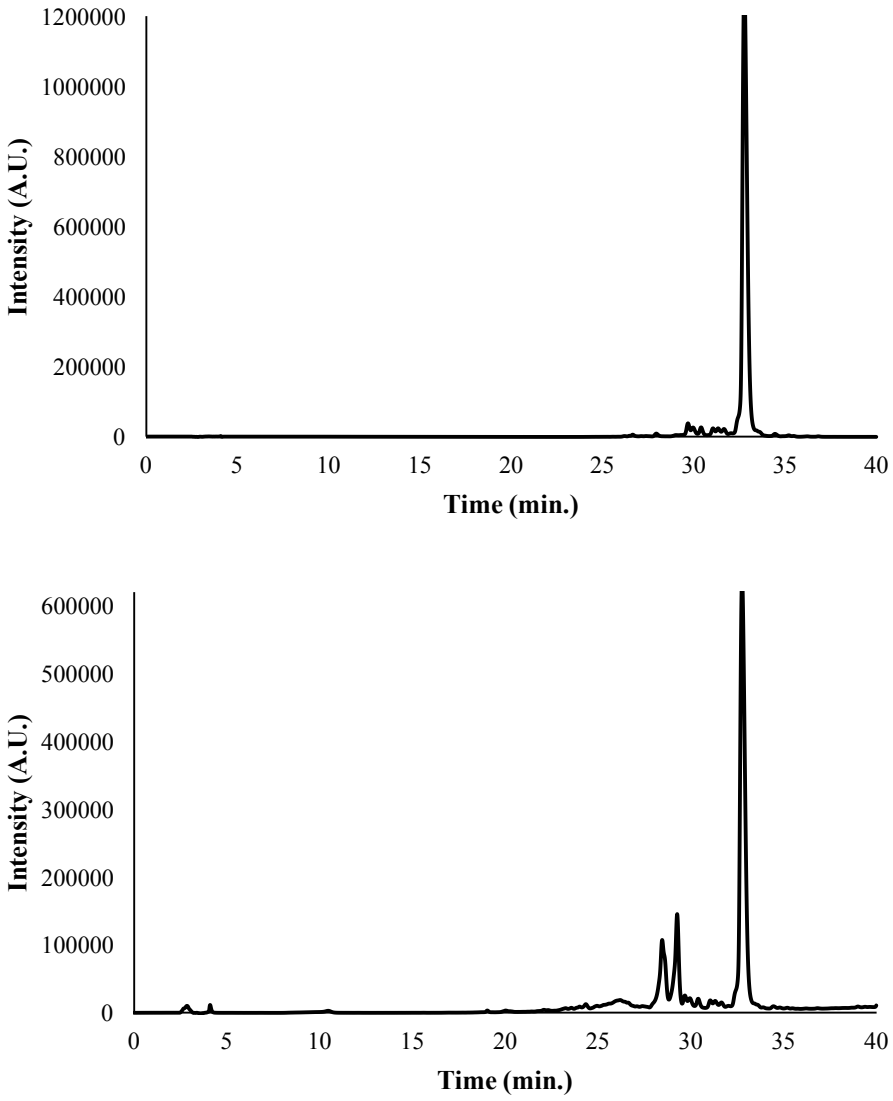
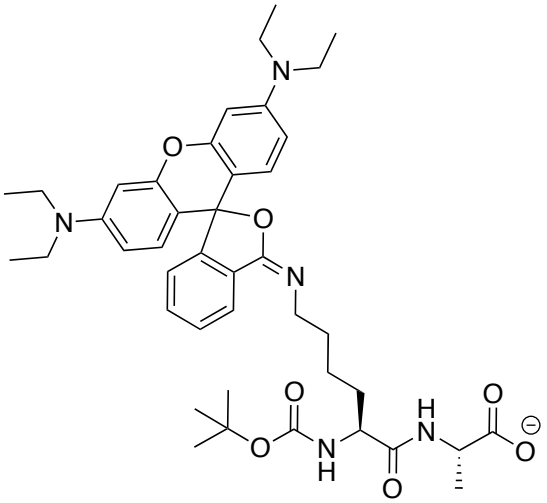
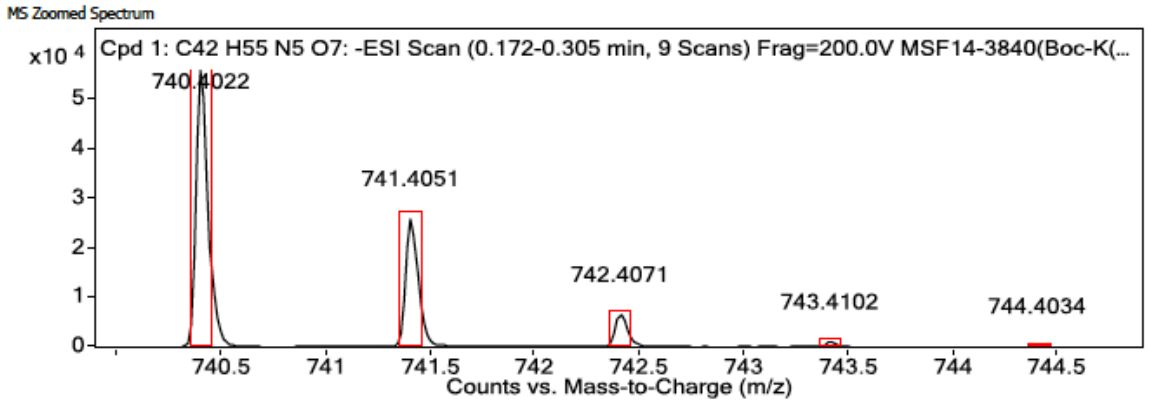


Figure 3.16: (top) Analytical trace for peptide 3.1 at 254 nm; Retention time: 32.736 min. (bottom) Analytical trace at 550 nm: Retention time: 32.74667 min.



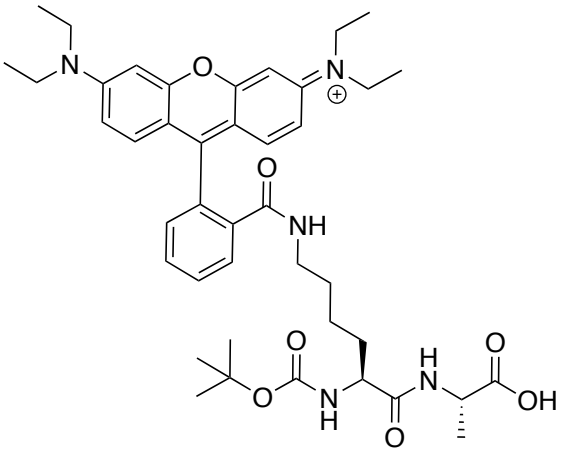
m/z: 740.40 (100.0%), 741.41 (45.4%), 742.41 (10.1%), 741.40 (1.8%), 742.41 (1.4%)
boc-lys(TMR)-ala



MS Spectrum Peak List

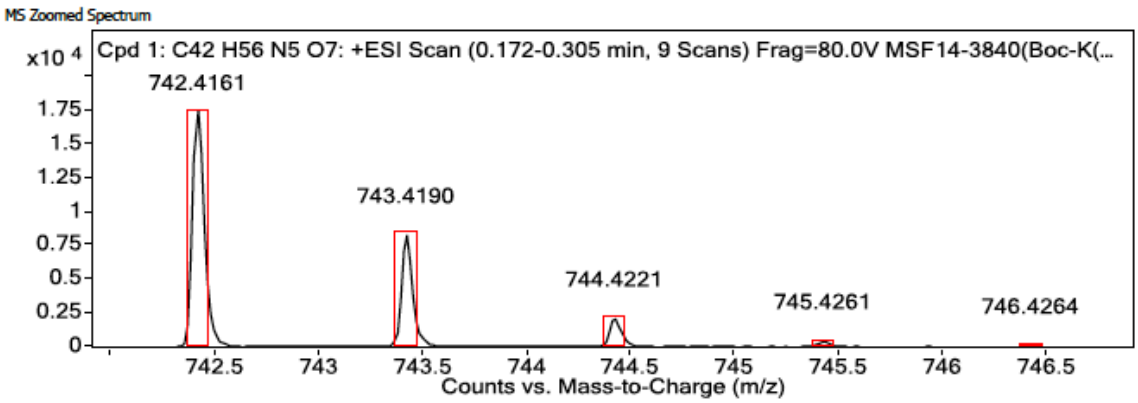
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
740.40220	740.40290	1	56343.59	C42H55N5O7	(M-H)-	0.97
741.40510	741.40600	1	26084.4	C42H55N5O7	(M-H)-	1.21
742.40710	742.40900	1	6694.42	C42H55N5O7	(M-H)-	2.5
743.41020	743.41170	1	1180.21	C42H55N5O7	(M-H)-	2.02
744.40340	744.41450	1	285.42	C42H55N5O7	(M-H)-	14.87

Figure 3.17: HRMS data for peptide 3.2. Starting amount: (0.12 mmole). Yield: 4%.



m/z: 742.42 (100.0%), 743.42 (45.4%), 744.42 (10.1%), 743.41 (1.8%), 744.42 (1.4%)

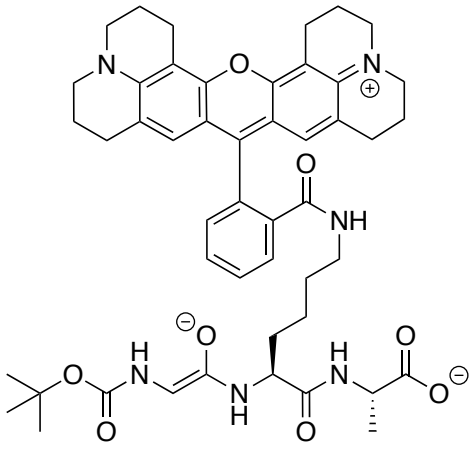
boc-lys(RhB_)-ala



MS Spectrum Peak List

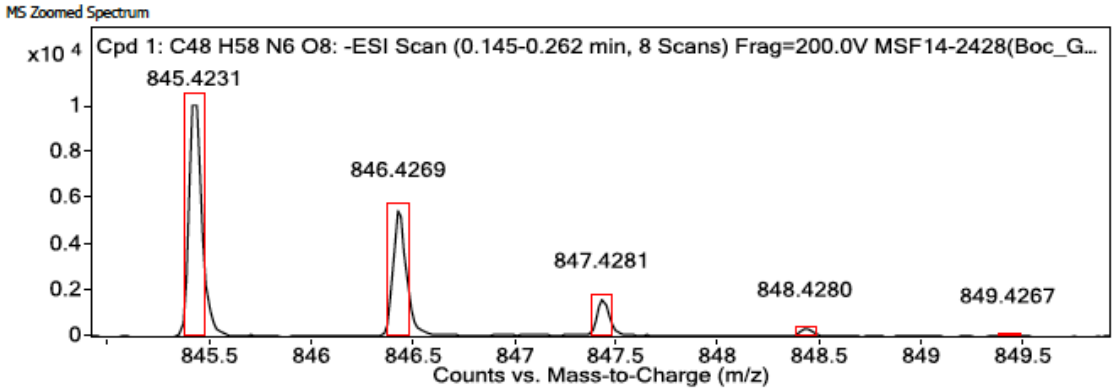
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
742.41610	742.41740	1	17535.91	C42H56N5O7	M+	1.8
743.41900	743.42060	1	8291.91	C42H56N5O7	M+	2.14
744.42210	744.42350	1	2105.96	C42H56N5O7	M+	1.84
745.42610	745.42630	1	440.99	C42H56N5O7	M+	0.26
746.42640	746.42900	1	104.79	C42H56N5O7	M+	3.47

Figure 3. 18: HRMS data for peptide 3.2.



m/z: 845.42 (100.0%), 846.43 (51.9%), 847.43 (13.2%), 846.42 (2.2%), 847.43 (1.6%), 848.43 (1.4%), 847.42 (1.2%)

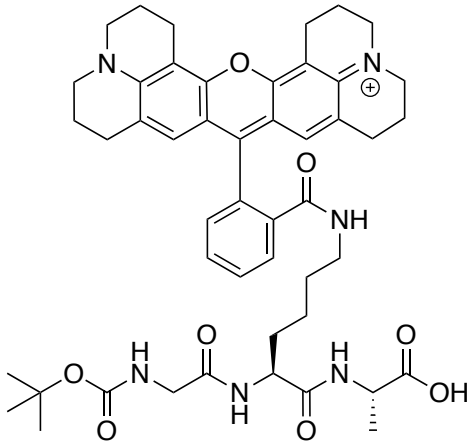
boc-gly-lys-ala



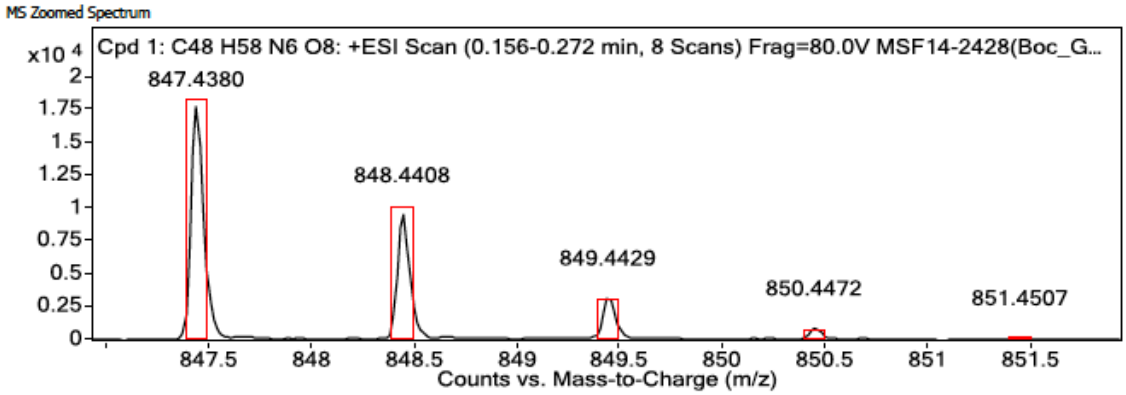
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
144.96400			47390.64			
845.42310	845.42430	1	10512.44	C48H58N6O8	(M-H)-	1.43
846.42690	846.42750	1	5601.52	C48H58N6O8	(M-H)-	0.71
847.42810	847.43040	1	1643.58	C48H58N6O8	(M-H)-	2.74
848.42800	848.43320	1	378.12	C48H58N6O8	(M-H)-	6.17
849.42670	849.43600	1	101.47	C48H58N6O8	(M-H)-	10.9

Figure 3. 19: HRMS data for peptide 3.3. Starting amount: 0.13 mmole. Yield: 3%



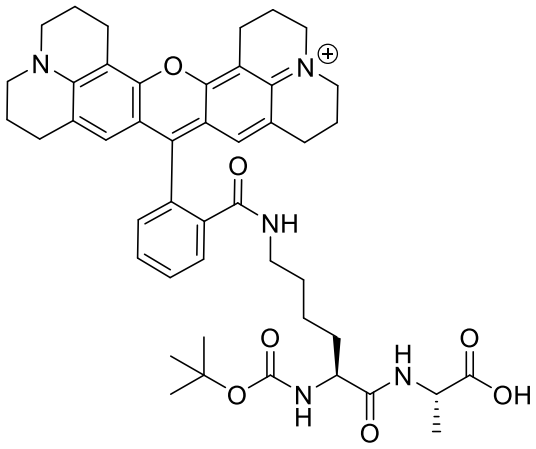
m/z: 847.44 (100.0%), 848.44 (51.9%), 849.45 (13.2%), 848.44 (2.2%), 849.44 (1.6%), 850.45 (1.4%), 849.44 (1.2%)



MS Spectrum Peak List

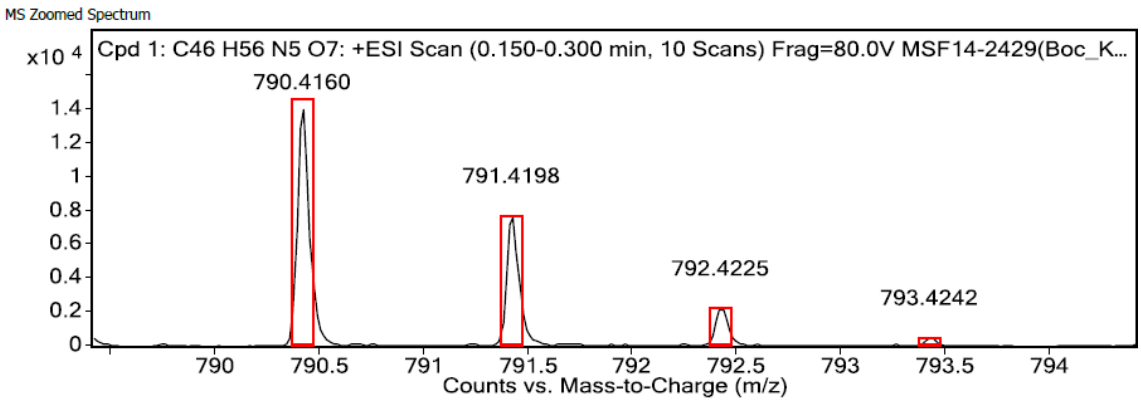
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
528.51090			35447.11			
847.43800	847.43890	1	17747.02	C48H58N6O8	(M+H)+	1.07
848.44080	848.44200	1	9594.32	C48H58N6O8	(M+H)+	1.41
849.44290	849.44500	1	3288.94	C48H58N6O8	(M+H)+	2.49
850.44720	850.44780	1	995.61	C48H58N6O8	(M+H)+	0.69
851.45070	851.45050	1	256.38	C48H58N6O8	(M+H)+	-0.2

Figure 3. 20: HRMS data for peptide 3.3.



m/z: 790.42 (100.0%), 791.42 (49.8%), 792.42 (12.1%), 791.41 (1.8%), 792.42 (1.4%), 793.43 (1.1%)

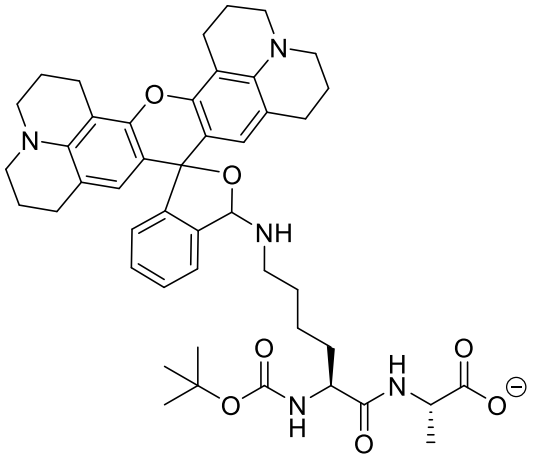
boc-lys(Rh101)-ala



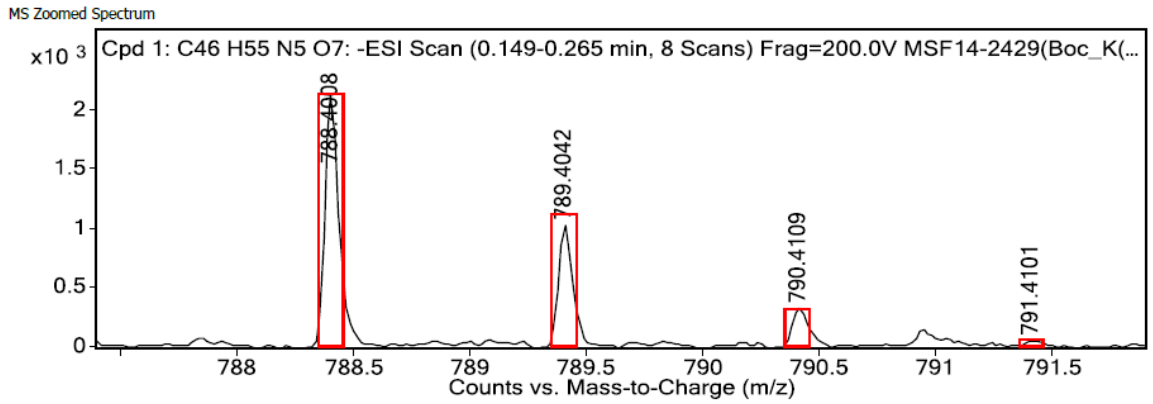
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
685.43460			18023.03			
790.41600	790.41740	1	14166.61	C46H56N5O7	M+	1.75
791.41980	791.42060	1	7780.39	C46H56N5O7	M+	0.98
792.42250	792.42360	1	2289.74	C46H56N5O7	M+	1.39
793.42420	793.42640	1	546.61	C46H56N5O7	M+	2.78
794.41950	794.42920	1	137.96	C46H56N5O7	M+	12.22

Figure 3.21 : HRMS data for peptide 3.4. Starting amount: (0.10 mmole). Yield: 3%.



m/z: 788.40 (100.0%), 789.41 (49.8%), 790.41 (12.1%), 789.40 (1.8%), 790.41 (1.4%), 791.41 (1.1%)



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
141.01610			11425.25			
788.40080	788.40290	1	2126.13	C46H55N5O7	(M-H)-	2.62
789.40420	789.40600	1	1039.68	C46H55N5O7	(M-H)-	2.39
790.41090	790.40900	1	337.4	C46H55N5O7	(M-H)-	-2.36
791.41010	791.41190	1	65.14	C46H55N5O7	(M-H)-	2.27

Figure 3. 22: HRMS data for peptide 3.4.

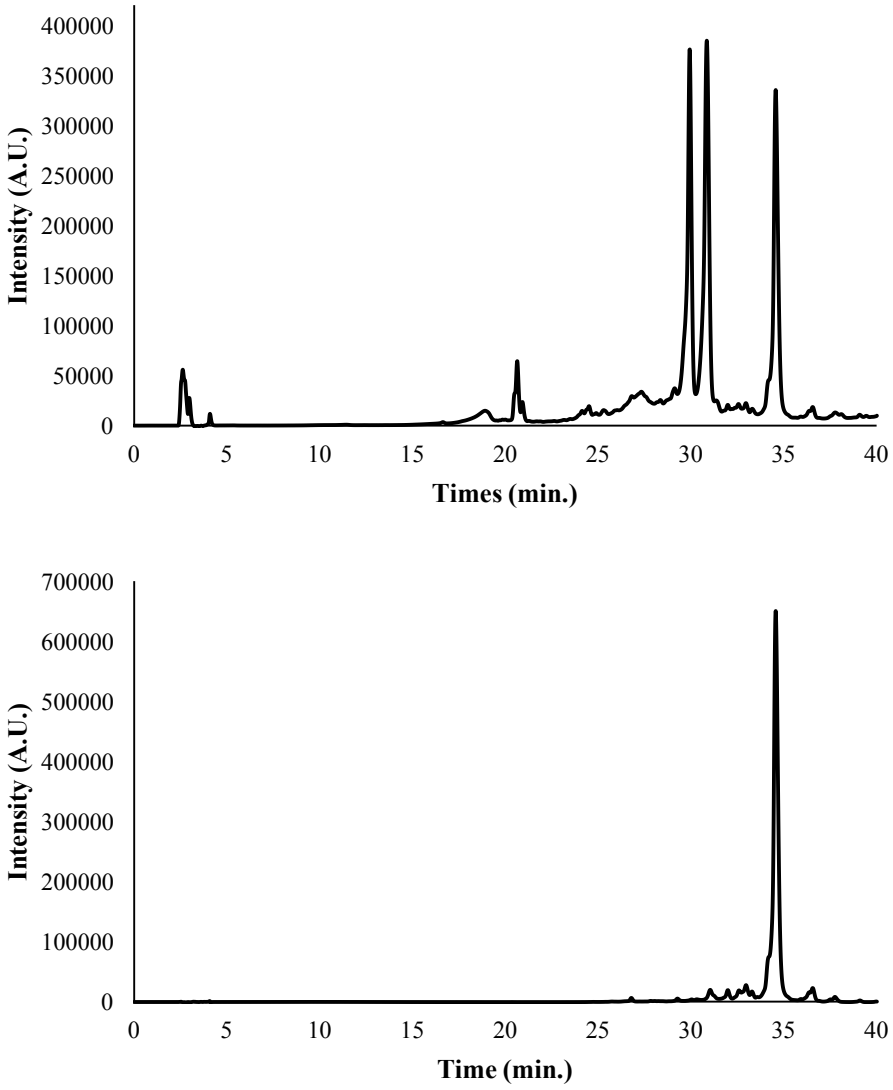
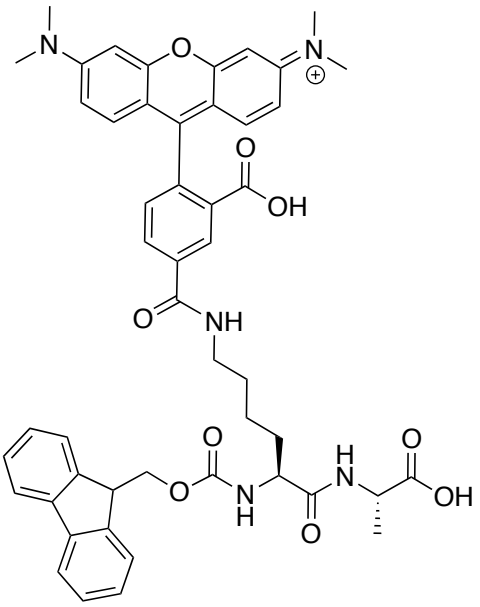
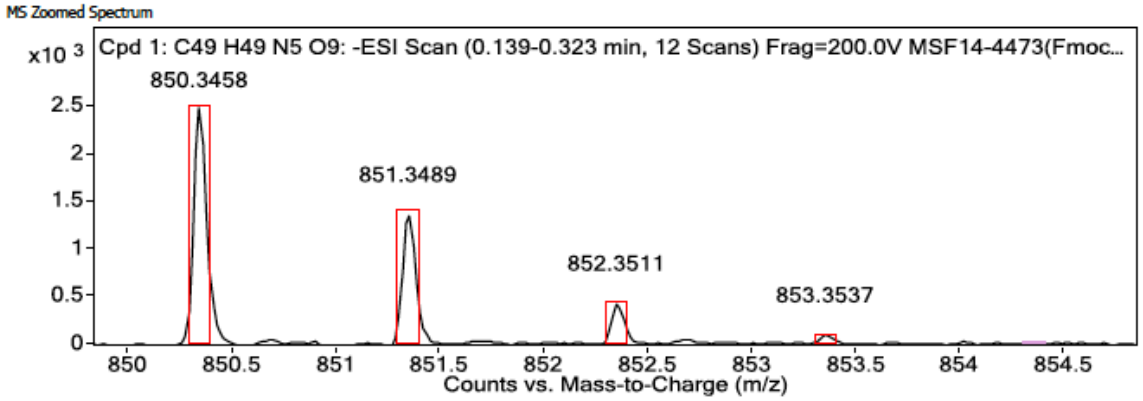


Figure 3.23: (top) Analytical trace for peptide 3.4 at 254 nm; Retention time: 29.92 min., 30.82667 min., 34.53867 min. (bottom) Analytical trace at 550 nm; Retention: 34.54933.



m/z: 852.36 (100.0%), 853.36 (53.0%), 854.37 (13.8%), 854.36 (1.8%), 853.36 (1.8%), 855.37 (1.5%)

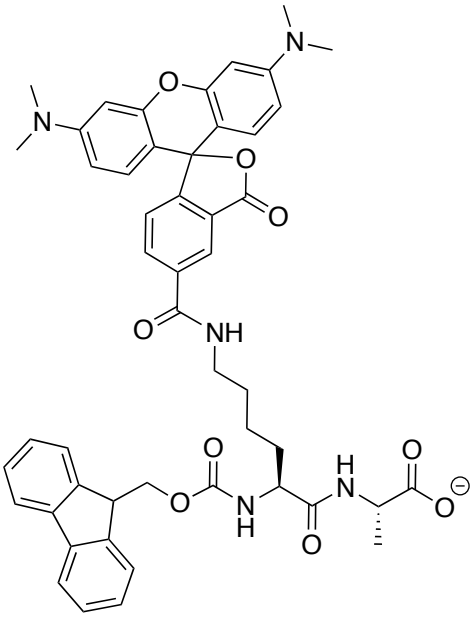
Fmoc-lys(Rh)-ala



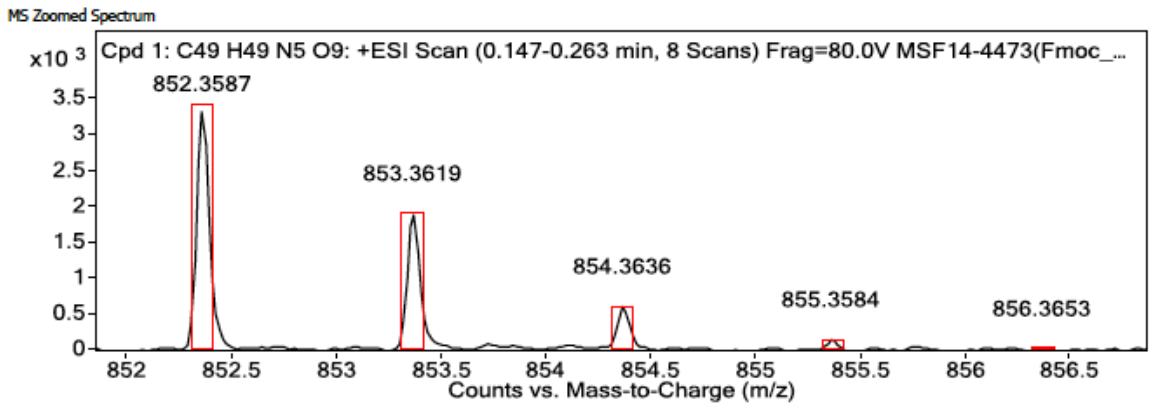
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
850.34580	850.34580	1	2489.34	C49H49N5O9	(M-H)-	-0.08
851.34890	851.34890	1	1373.4	C49H49N5O9	(M-H)-	0
852.35110	852.35190	1	441.7	C49H49N5O9	(M-H)-	0.93
853.35370	853.35470	1	108.59	C49H49N5O9	(M-H)-	1.14

Figure 3.24: HRMS data for peptide 3.5. Starting amount: (0.10 mmole). Yield: 62%.



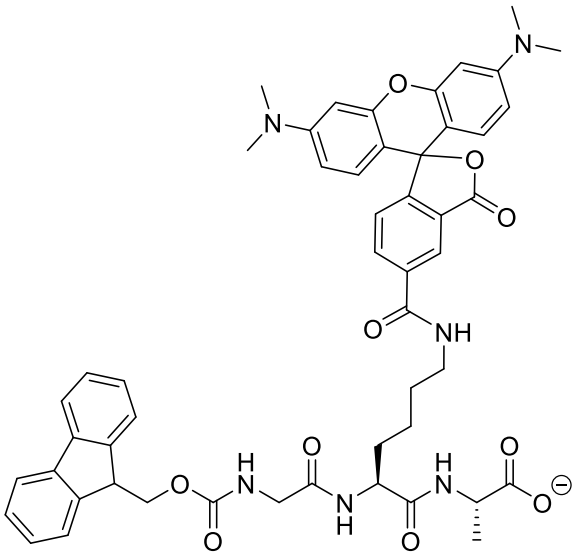
m/z: 850.35 (100.0%), 851.35 (53.0%), 852.35 (13.8%),852.35 (1.8%), 851.34 (1.8%), 853.36 (1.5%)



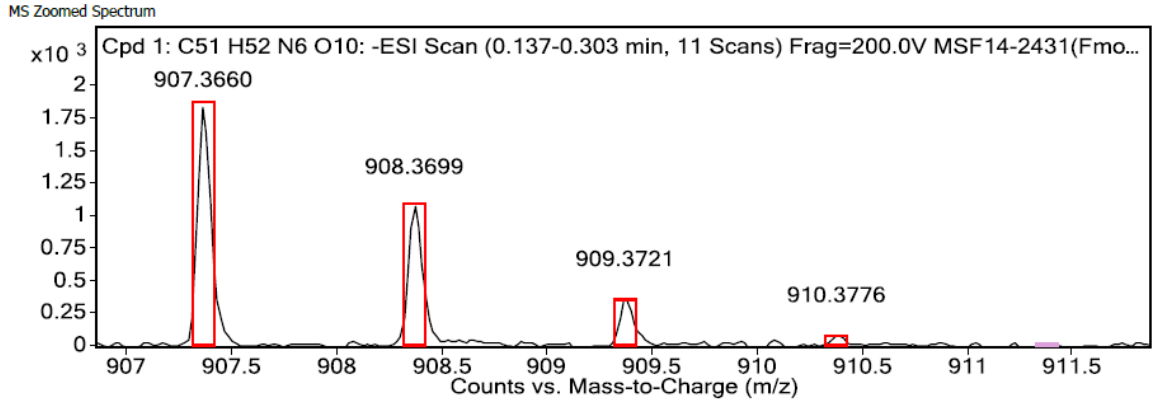
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
360.32340			30848.13			
852.35870	852.36030	1	3338.52	C49H49N5O9	(M+H)+	1.85
853.36190	853.36350	1	1916.95	C49H49N5O9	(M+H)+	1.89
854.36360	854.36640	1	619.85	C49H49N5O9	(M+H)+	3.37
855.35840	855.36930	1	156.64	C49H49N5O9	(M+H)+	12.66
856.36530	856.37200	1	28.88	C49H49N5O9	(M+H)+	7.86

Figure 3.25: HRMS data for peptide 3.5.



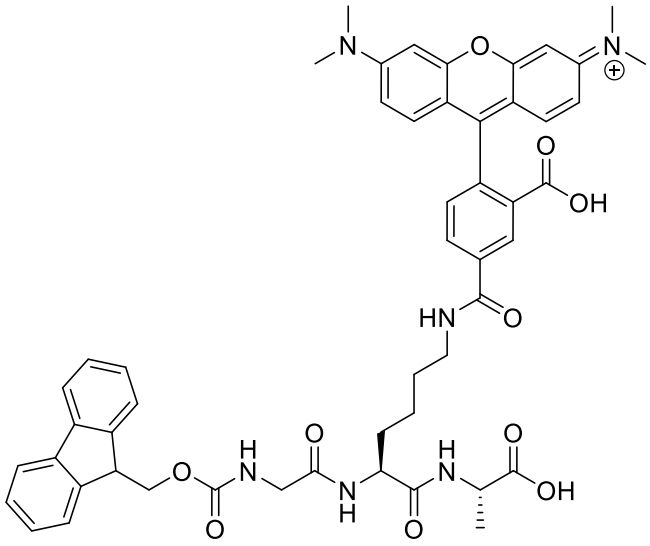
m/z: 907.37 (100.0%), 908.37 (55.2%), 909.37 (14.9%), 908.36 (2.2%), 909.37 (2.1%), 910.38 (1.8%), 909.37 (1.2%), 910.37 (1.1%)



MS Spectrum Peak List

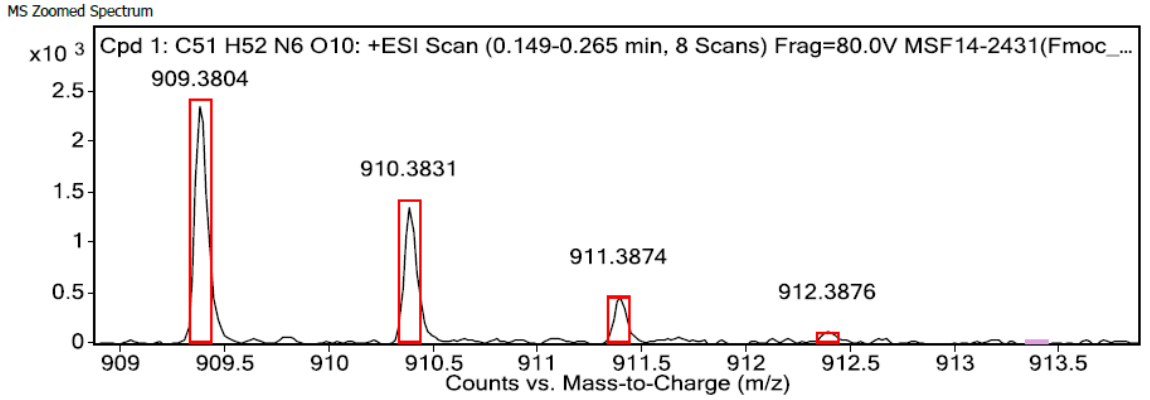
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
907.36600	907.36720	1	1851.36	C51H52N6O10	(M-H)-	1.38
908.36990	908.37040	1	1079.63	C51H52N6O10	(M-H)-	0.56
909.37210	909.37330	1	369.9	C51H52N6O10	(M-H)-	1.32
910.37760	910.37610	1	92.53	C51H52N6O10	(M-H)-	-1.64
1033.98670			2425.81			

Figure 3.26: HRMS data for peptide 3.6. Starting amount: 0.10 mmole. Yield: 34%.



m/z: 909.38 (100.0%), 910.39 (55.2%), 911.39 (14.9%), 912.39 (2.6%), 910.38 (2.2%), 911.39 (2.1%), 911.38 (1.2%), 912.39 (1.1%)

Fmoc-gly-lys(TMR)-ala



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
360.32340			24304.56			
909.38040	909.38180	1	2402.64	C51H52N6O10	(M+H)+	1.46
910.38310	910.38490	1	1345.82	C51H52N6O10	(M+H)+	2.05
911.38740	911.38780	1	473.04	C51H52N6O10	(M+H)+	0.51
912.38760	912.39060	1	123.93	C51H52N6O10	(M+H)+	3.33

Figure 3.27: HRMS data for peptide 3.6.

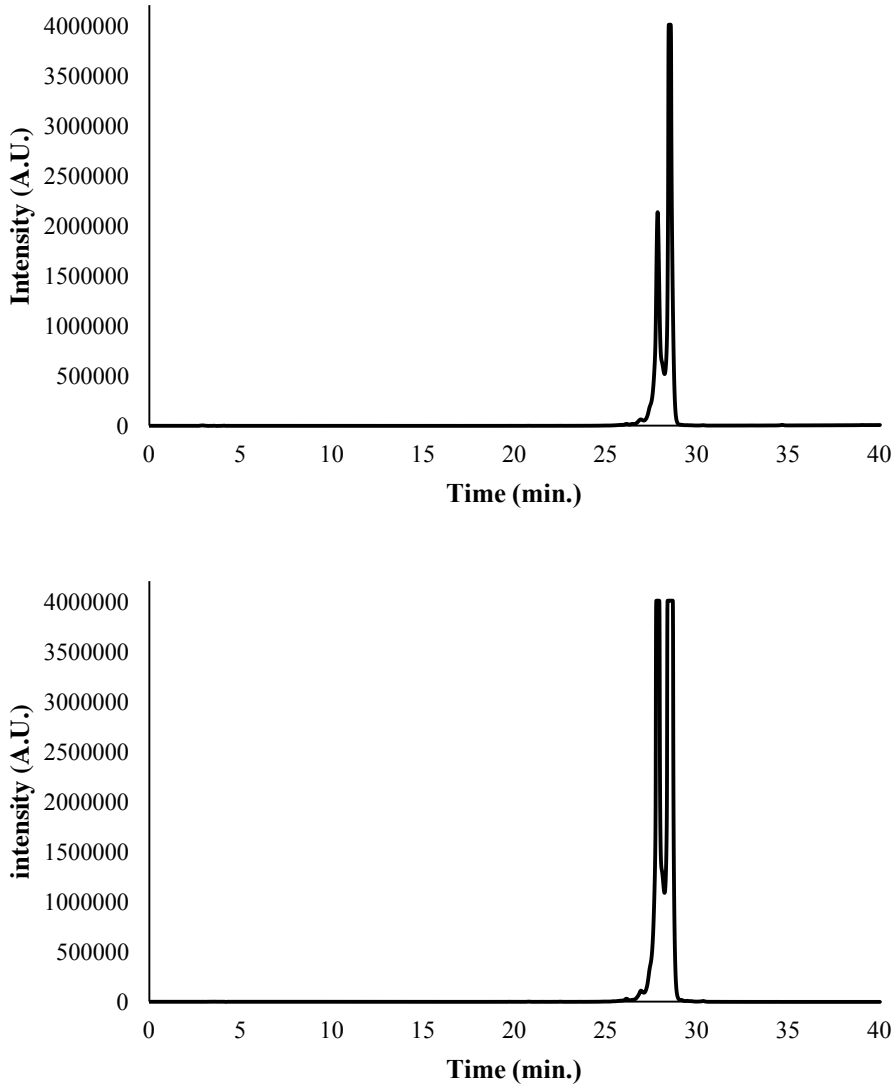
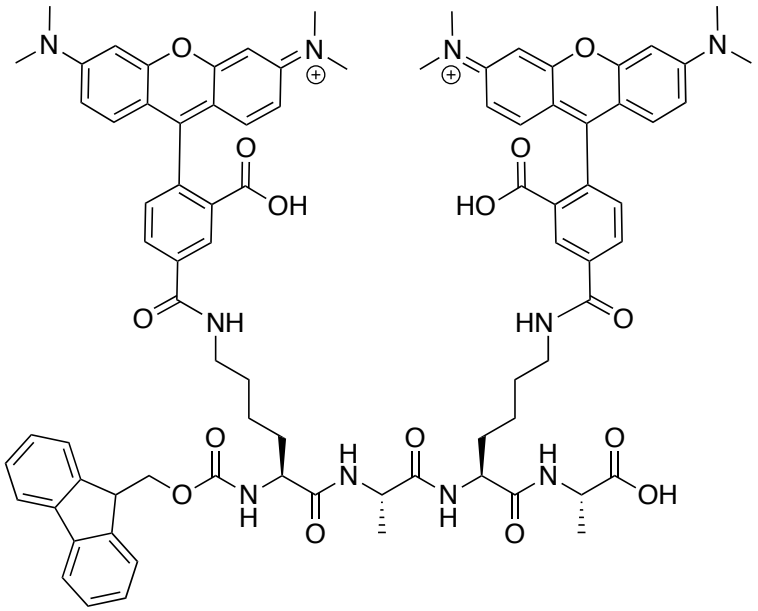
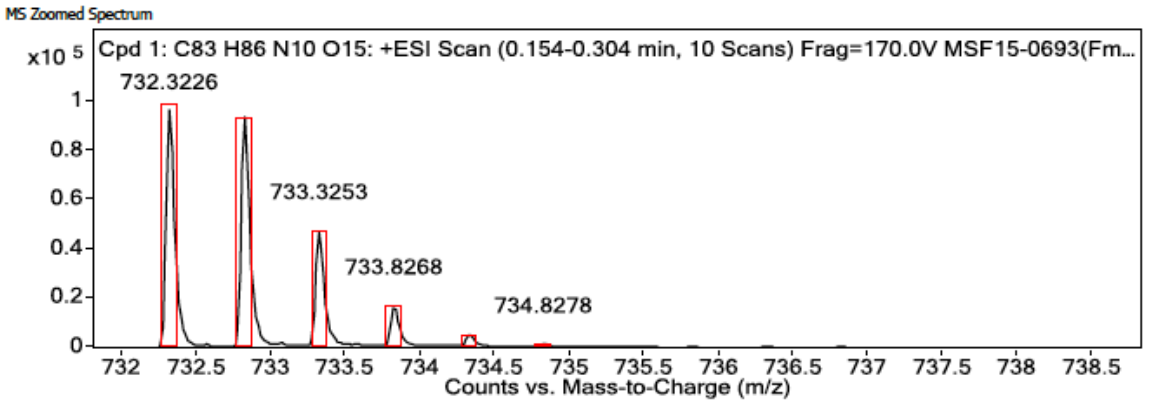


Figure 3.28: (top) Analytical trace for peptide 3.6 at 254 nm: Retention times: 27.81867 min., 28.448 min. (bottom) Analytical trace at 550 nm; Retention time: 27.76533 min., 27.88267 min.



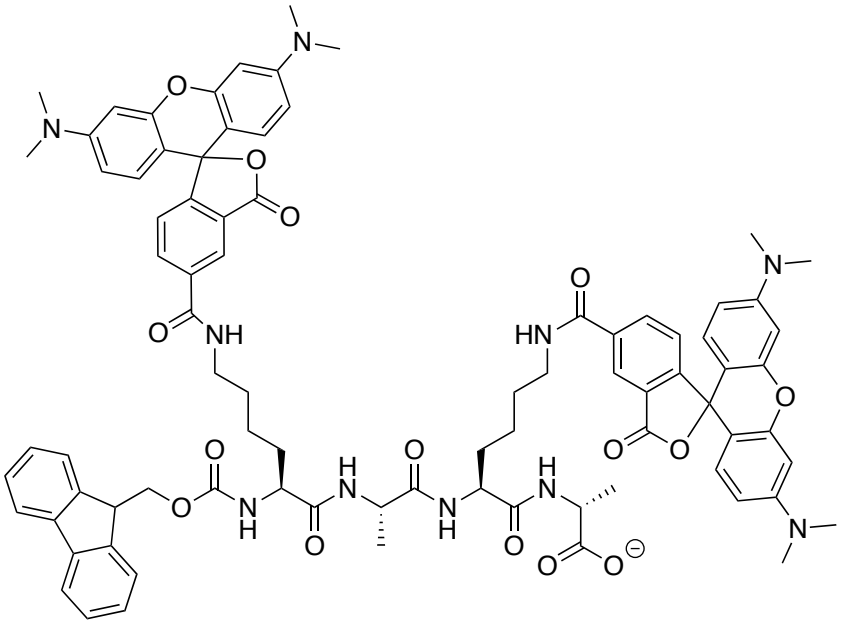
m/z: 732.32 (100.0%), 732.82 (89.8%), 733.32 (39.8%), 733.83 (10.8%), 732.82 (3.7%), 733.32 (3.3%), 733.32 (3.1%), 733.82 (2.8%), 734.33 (2.1%), 733.82 (1.5%), 734.33 (1.2%)



MS Spectrum Peak List

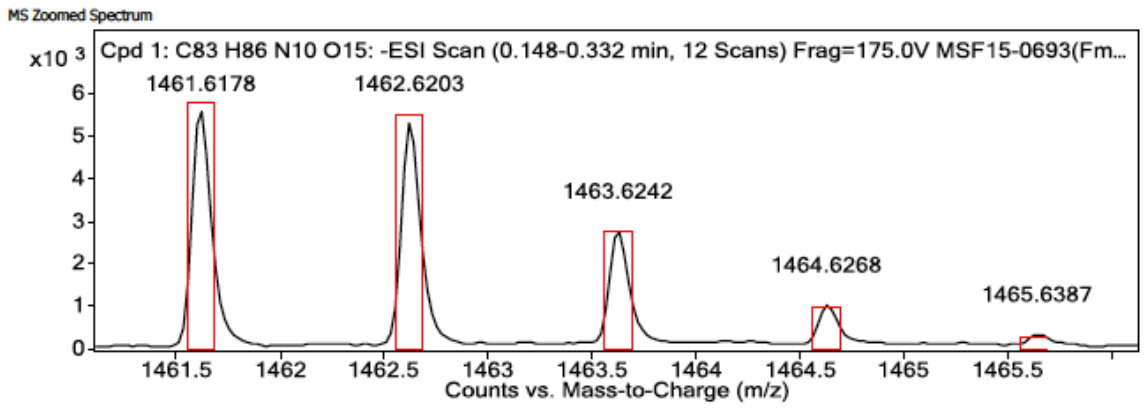
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
174.96900			111334.4			
732.32260	732.32100	2	96655.57	C83H86N10O15	(M+2H)+2	-2.14
732.82400	732.82260	2	93883.96	C83H86N10O15	(M+2H)+2	-1.95
733.32530	733.32410	2	47231.28	C83H86N10O15	(M+2H)+2	-1.7
733.82680	733.82550	2	16689.38	C83H86N10O15	(M+2H)+2	-1.72
734.32770	734.32690	2	4601.44	C83H86N10O15	(M+2H)+2	-1.01
734.82780	734.82830	2	1057.14	C83H86N10O15	(M+2H)+2	0.76
735.32770	735.32970	2	287.92	C83H86N10O15	(M+2H)+2	2.76

Figure 3.29: HRMS data for peptide 3.8. Starting amount: 0.035 mmole. Yield: 9%.



m/z: 1461.62 (100.0%), 1462.62 (89.8%), 1463.63 (39.8%), 1464.63 (10.8%), 1462.62 (3.7%), 1463.62 (3.3%), 1463.62 (3.1%), 1464.63 (2.8%), 1465.63 (2.1%), 1464.62 (1.5%), 1465.63 (1.2%)

Fmoc-lys(TMR)-ala-lys(TMR)-ala



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
980.01650			249724.94			
1461.61780	1461.62010	1	5669.21	C83H86N10O15	(M-H)-	1.6
1462.62030	1462.62330	1	5377.06	C83H86N10O15	(M-H)-	2.04
1463.62420	1463.62630	1	2809.91	C83H86N10O15	(M-H)-	1.46
1464.62680	1464.62920	1	1074.82	C83H86N10O15	(M-H)-	1.68
1465.63870	1465.63210	1	382.93	C83H86N10O15	(M-H)-	-4.55

Figure 3.30: HRMS data for peptide 3.8.

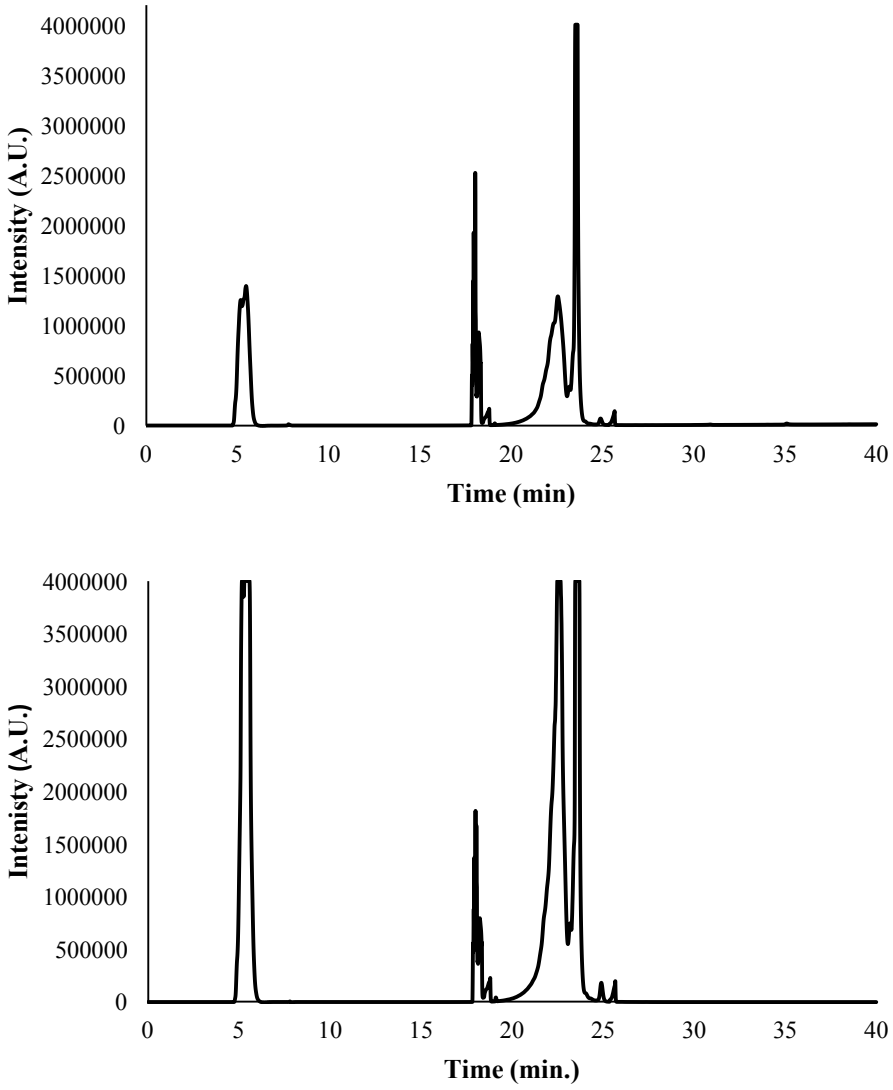
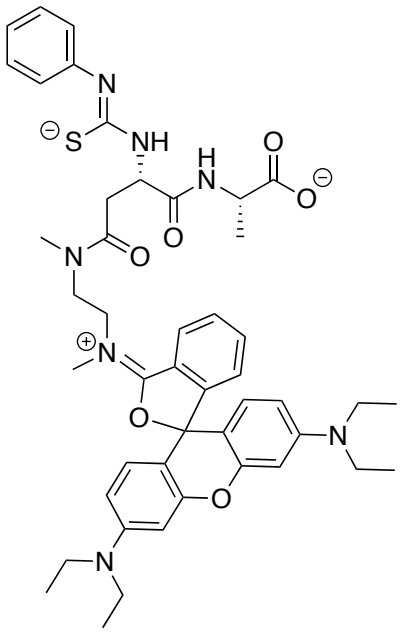
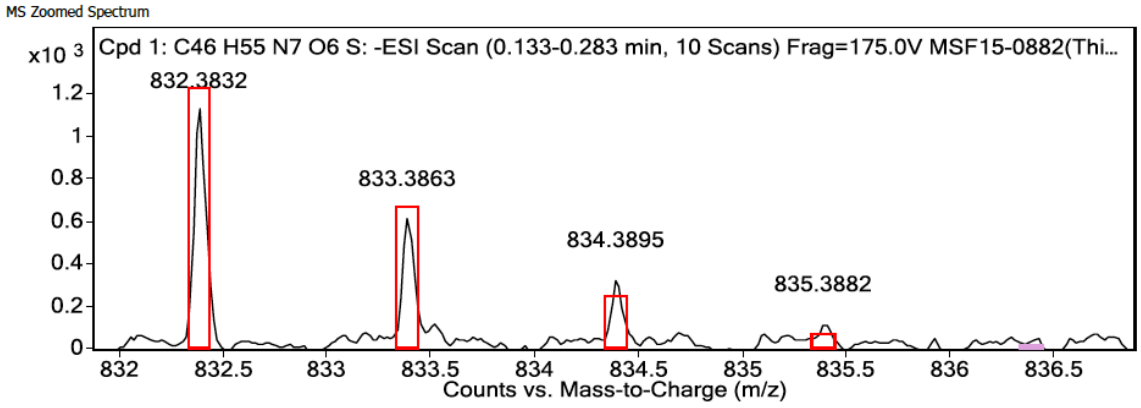


Figure 3.31: (top) Analytical trace for peptide 3.8 at 254 nm; Retention time: 5.10933 min., 18.03733 min., 23.52 min. (bottom) Analytical trace at 550 nm; Retention time: 5.152 min., 17.99467 min., 22.4767 min., 23.466 min.



m/z: 832.39 (100.0%), 833.39 (49.8%), 834.39 (12.1%), 834.38 (4.5%), 833.38 (2.6%), 835.39 (2.2%), 834.39 (1.3%), 834.39 (1.2%), 835.40 (1.1%)

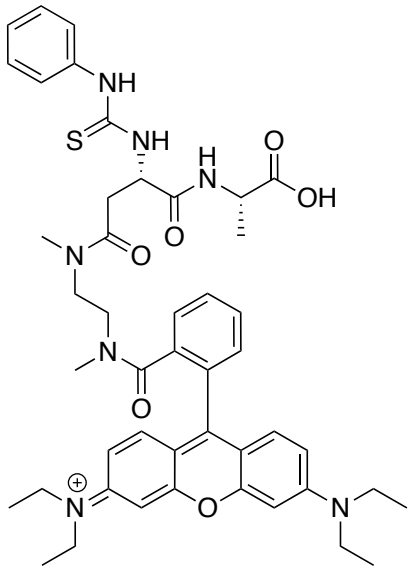
thiourea-asp(RhB_NNdimethylethylenediamine)-ala



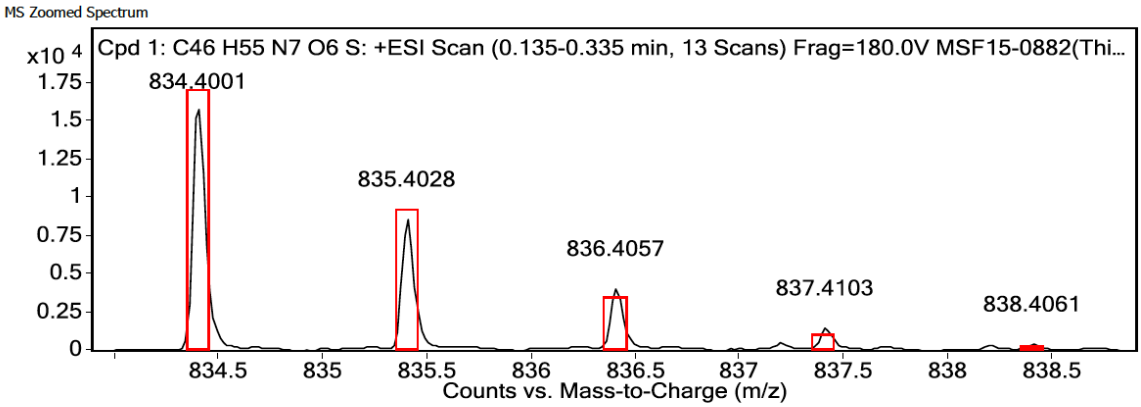
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
366.03680			12780.91			
832.38320	832.38620	1	1141.37	C46H55N7O6S	(M-H)-	3.61
833.38630	833.38920	1	616.79	C46H55N7O6S	(M-H)-	3.51
834.38950	834.38980	1	328.81	C46H55N7O6S	(M-H)-	0.36
835.38820	835.39060	1	120.38	C46H55N7O6S	(M-H)-	2.89

Figure 3.32: HRMS data for peptide 3.9. Starting amount: 0.012 mmole. Yield: 14%.



m/z: 416.70 (100.0%), 417.20 (49.8%), 417.70 (12.1%), 417.69 (4.5%), 417.19 (2.6%), 418.20 (2.2%), 417.70 (1.3%), 417.70 (1.2%), 418.20 (1.1%)



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
360.32390			27325.87			
834.40010	834.40070	1	16132.64	C46H55N7O6S	(M+H)+	0.75
835.40280	835.40380	1	8610.89	C46H55N7O6S	(M+H)+	1.2
836.40570	836.40440	1	4067.27	C46H55N7O6S	(M+H)+	-1.6
837.41030	837.40520	1	1485.02	C46H55N7O6S	(M+H)+	-6.12
838.40610	838.40650	1	430.19	C46H55N7O6S	(M+H)+	0.45

Figure 3.33: HRMS data for peptide 3.10.

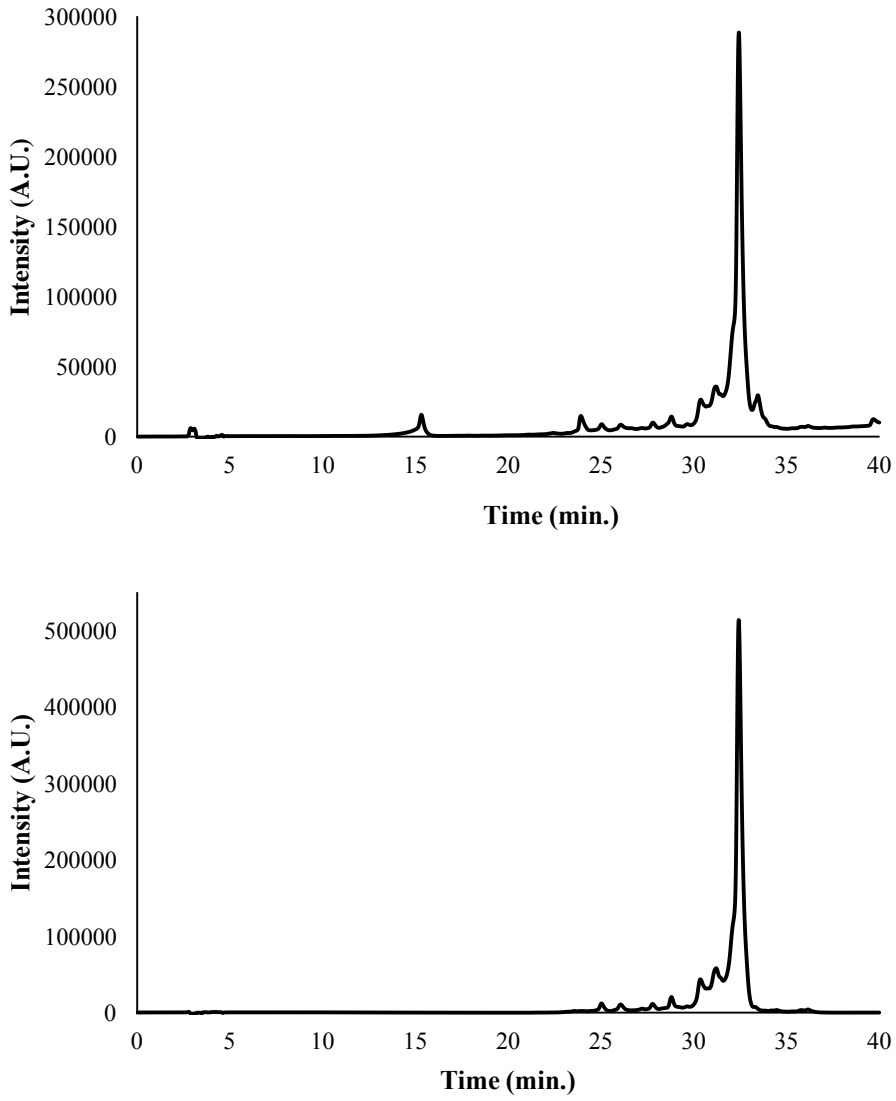
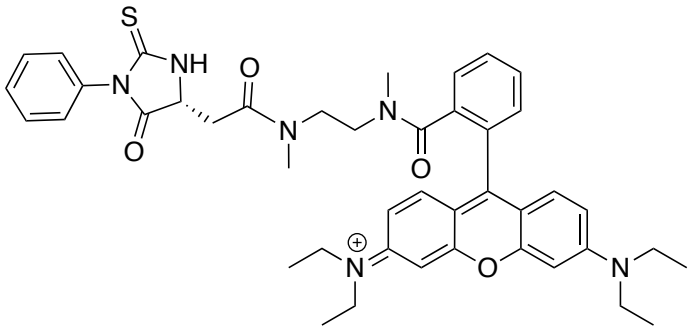
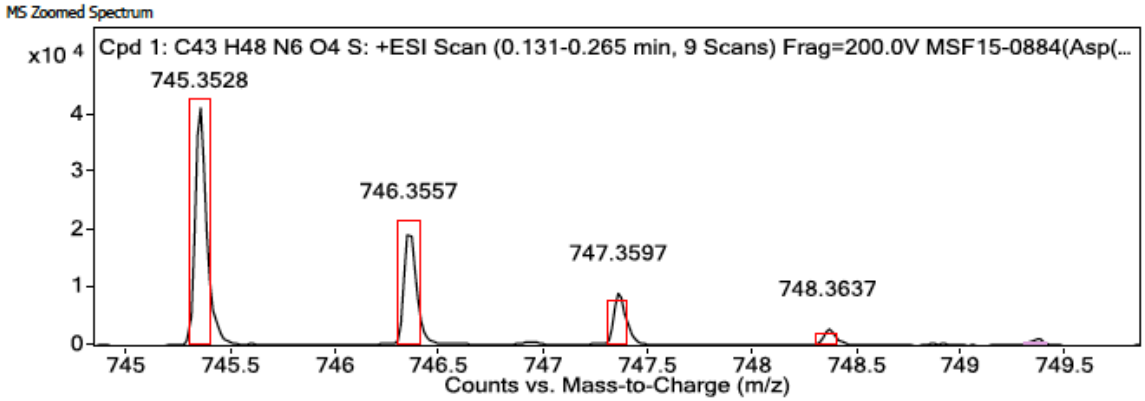


Figure 3.34: (top) Analytical trace for peptide 3.10 at 254 nm; Retention time: 32.416 min. (bottom) Analytical trace at 550 nm; Retention time: 32.416 min.



m/z: 745.35 (100.0%), 746.36 (46.5%), 747.36 (10.6%), 747.35 (4.5%), 746.35 (2.2%), 748.35 (2.1%), 747.35 (1.0%)



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
304.26140			134441.92			
745.35280	745.35310	1	41406.67	C43H48N6O4S	(M+H)+	0.3
746.35570	746.35610	1	19956.8	C43H48N6O4S	(M+H)+	0.58
747.35970	747.35640	1	9223.97	C43H48N6O4S	(M+H)+	-4.35
748.36370	748.35710	1	3026.93	C43H48N6O4S	(M+H)+	-8.76

Figure 3.35: HRMS data for peptide 3.11.

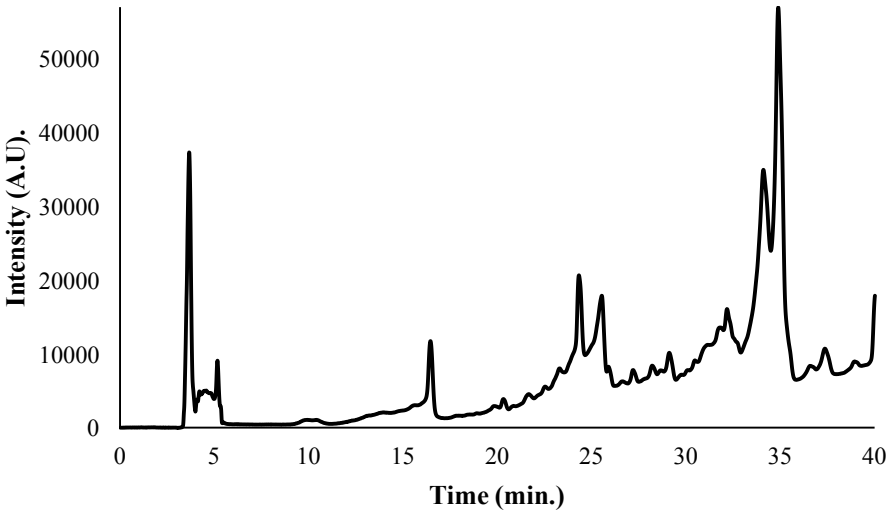
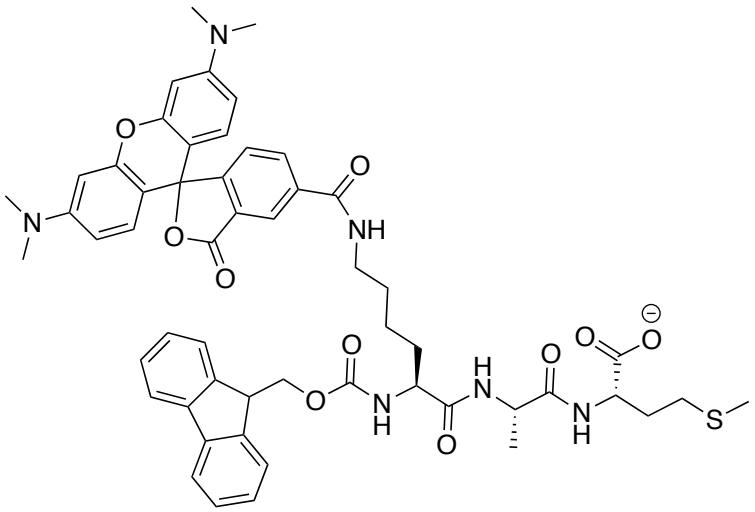
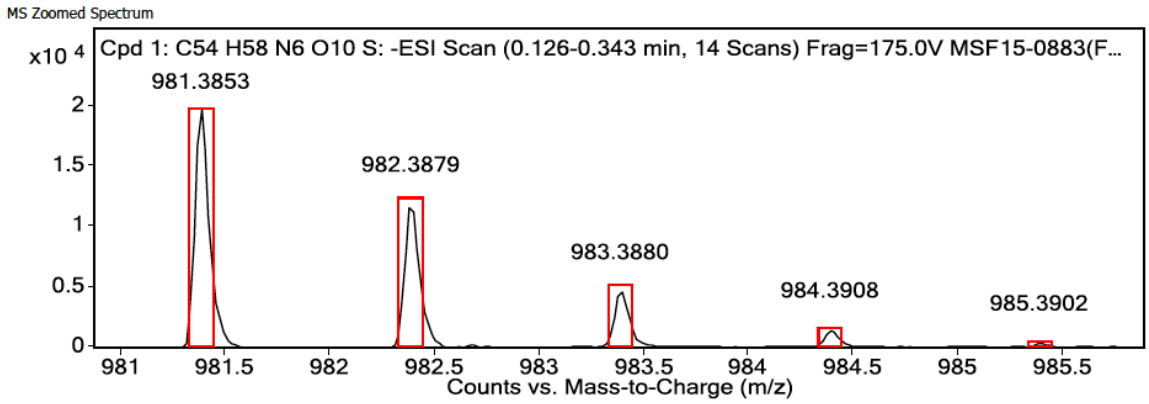


Figure 3.36: (top) Analytical trace for peptide 3.11 at 254 nm; Retention time: 34.85867 min.



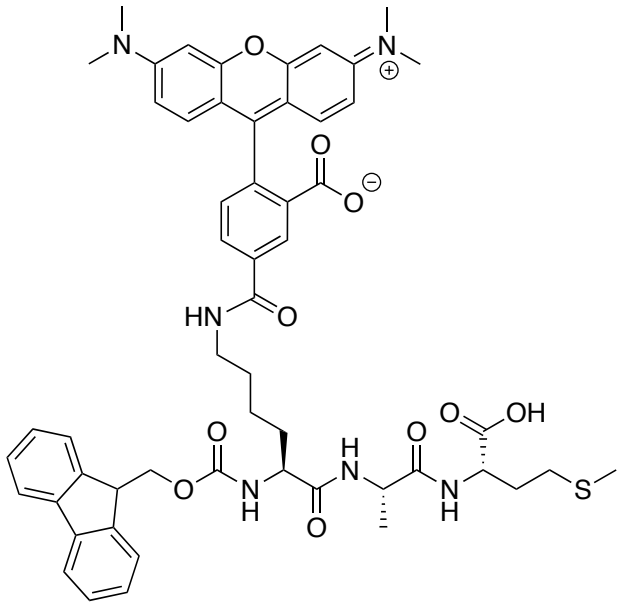
m/z: 981.39 (100.0%), 982.39 (58.4%), 983.39 (16.7%), 983.38 (4.5%), 984.39 (2.6%), 984.40 (2.3%), 982.38 (2.2%), 983.39 (2.1%), 983.39 (1.3%), 984.39 (1.2%)



MS Spectrum Peak List

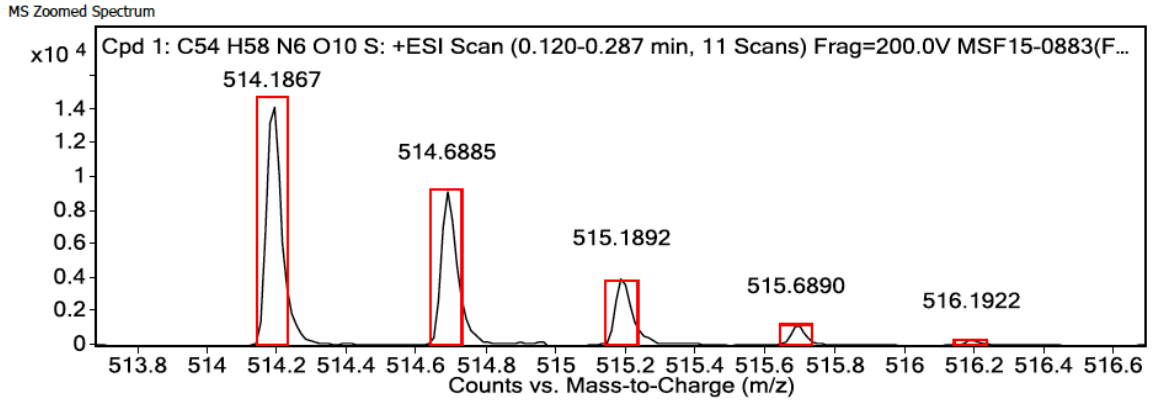
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
981.38530	981.38620	1	19762.9	C54H58N6O10S	(M-H)-	0.98
982.38790	982.38940	1	11890.9	C54H58N6O10S	(M-H)-	1.52
983.38800	983.39050	1	4634.58	C54H58N6O10S	(M-H)-	2.57
984.39080	984.39160	1	1451.29	C54H58N6O10S	(M-H)-	0.87
985.39020	985.39310	1	378.74	C54H58N6O10S	(M-H)-	2.9

Figure 3.37: HRMS data for peptide 3.12. Starting amount: 0.025 mmole. Yield: 55%.



m/z: 514.19 (100.0%), 514.69 (58.4%), 515.19 (16.7%), 515.18 (4.5%), 515.69 (2.6%), 515.69 (2.3%), 514.68 (2.2%), 515.19 (2.1%), 515.19 (1.3%), 515.69 (1.2%)

Fmoc-lys(TMR)-ala-met



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
503.19600			24928.17			
514.18670	514.18600	2	14412.29	C54H58N6O10S	(M+2Na)+2	-1.38
514.68850	514.68750	2	9175.31	C54H58N6O10S	(M+2Na)+2	-1.79
515.18920	515.18810	2	4069.84	C54H58N6O10S	(M+2Na)+2	-2.22
515.68900	515.68870	2	1216.1	C54H58N6O10S	(M+2Na)+2	-0.65
516.19220	516.18940	2	323.15	C54H58N6O10S	(M+2Na)+2	-5.46
516.68580	516.69020	2	80.33	C54H58N6O10S	(M+2Na)+2	8.6

Figure 3.38: HRMS data for peptide 3.12.

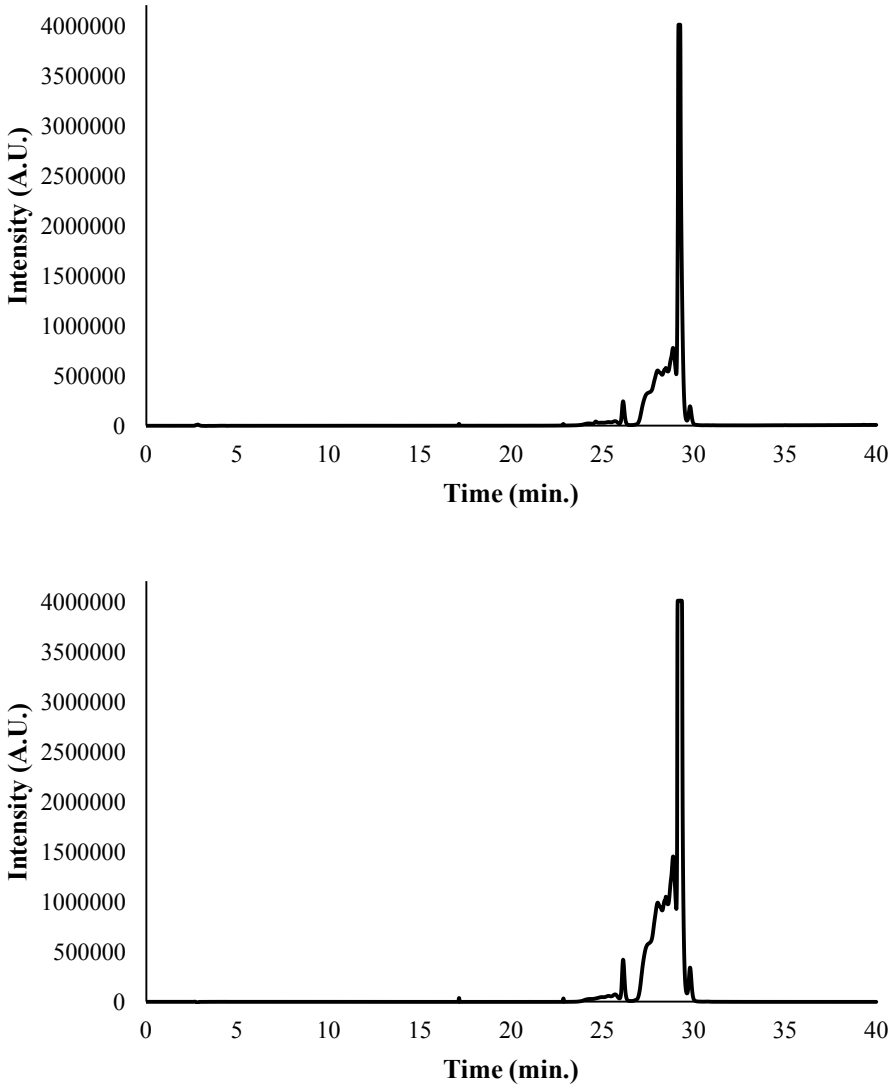
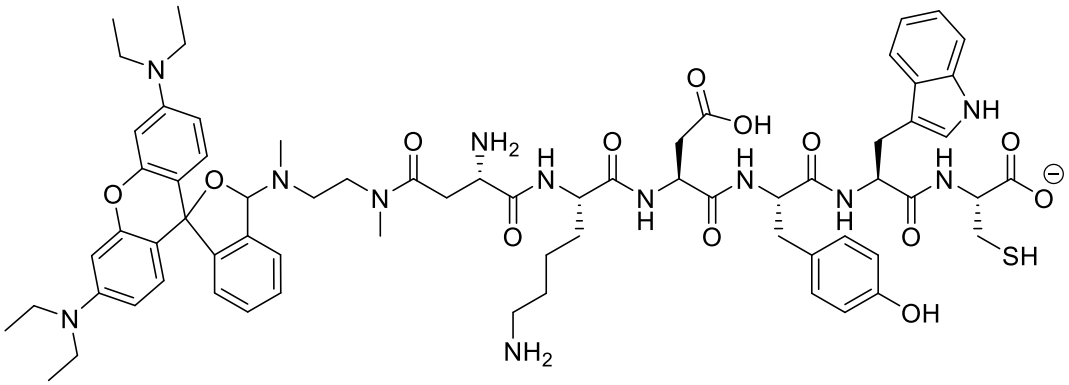
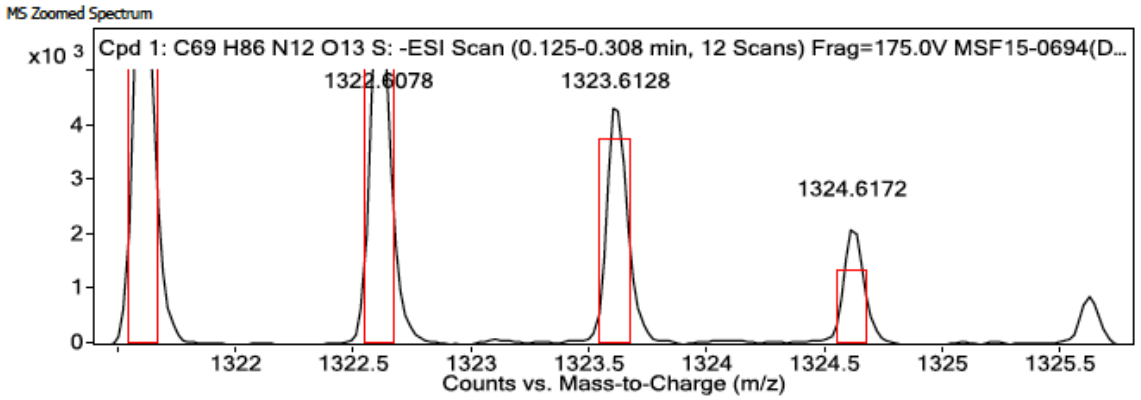


Figure 3.39: (top) Analytical trace for peptide 3.12 at 254 nm; Retention time: 29.16267 min. (bottom) Analytical trace at 550 nm; Retention time: 29.12 min.



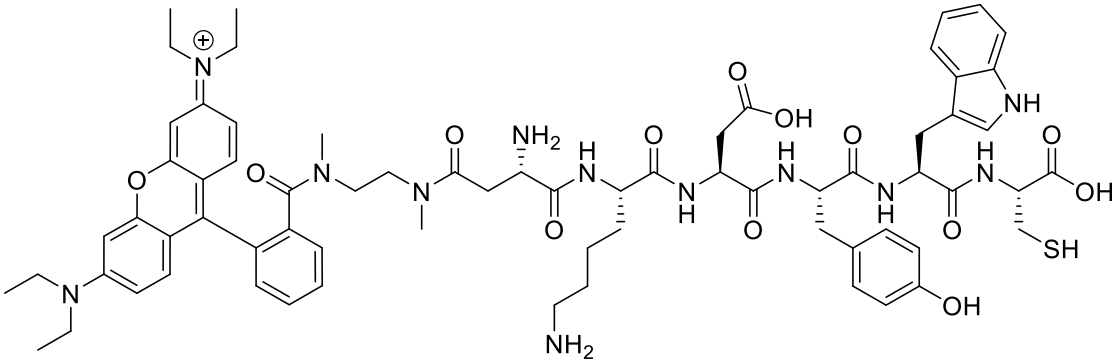
m/z: 1323.62 (100.0%), 1324.63 (74.6%), 1325.63 (27.4%), 1326.63 (6.6%), 1325.62 (4.5%), 1324.62 (4.1%), 1326.62 (3.4%), 1325.62 (3.3%), 1325.63 (2.7%), 1326.63 (2.0%), 1327.63 (1.2%), 1326.63 (1.2%), 1327.64 (1.2%), 1324.63 (1.0%)



MS Spectrum Peak List

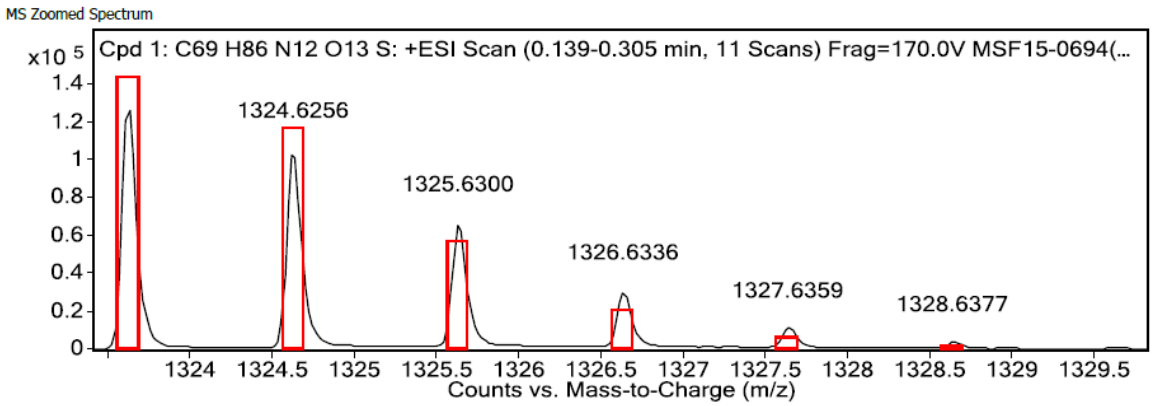
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
520.90980			1267027.21			
1321.60500	1321.60850	1	8218.6	C69H86N12O13S	(M-H)-	2.67
1322.60780	1322.61150	1	6685.26	C69H86N12O13S	(M-H)-	2.8
1323.61280	1323.61330	1	4466.64	C69H86N12O13S	(M-H)-	0.36
1324.61720	1324.61470	1	2135.49	C69H86N12O13S	(M-H)-	-1.83
1325.61720	1325.61630	1	875.91	C69H86N12O13S	(M-H)-	-0.66
1326.61120	1326.61800	1	228.56	C69H86N12O13S	(M-H)-	5.1

Figure 3.40: HRMS data for peptide 3.13. Starting amount: 0.23 mmole. Yield: 1.1%.



m/z: 1323.62 (100.0%), 1324.63 (74.6%), 1325.63 (27.4%), 1326.63 (6.6%), 1325.62 (4.5%), 1324.62 (4.1%), 1326.62 (3.4%), 1325.62 (3.3%), 1325.63 (2.7%), 1326.63 (2.0%), 1327.63 (1.2%), 1326.63 (1.2%), 1327.64 (1.2%), 1324.63 (1.0%)

asp(RhB)-lys-asp-tyr-trp-cys



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
662.31630			974077.42			
1323.62270	1323.62310	1	128311.72	C69H86N12O13S	(M+H)+	0.26
1324.62560	1324.62610	1	105694.62	C69H86N12O13S	(M+H)+	0.41
1325.63000	1325.62780	1	66749.12	C69H86N12O13S	(M+H)+	-1.66
1326.63360	1326.62930	1	30617.45	C69H86N12O13S	(M+H)+	-3.24
1327.63590	1327.63080	1	10470.73	C69H86N12O13S	(M+H)+	-3.81
1328.63770	1328.63250	1	3216.24	C69H86N12O13S	(M+H)+	-3.92
1329.63210	1329.63440	1	984.69	C69H86N12O13S	(M+H)+	1.66

Figure 3.41: HRMS data for peptide 3.13.

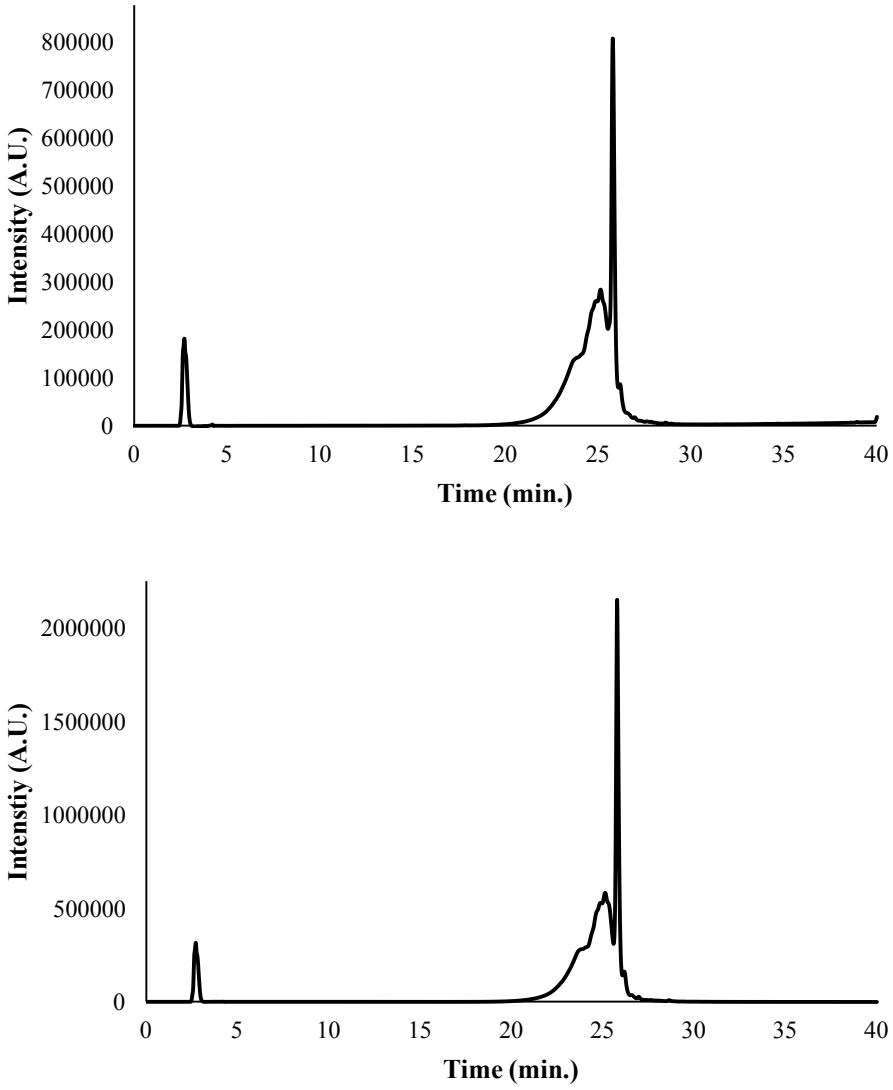
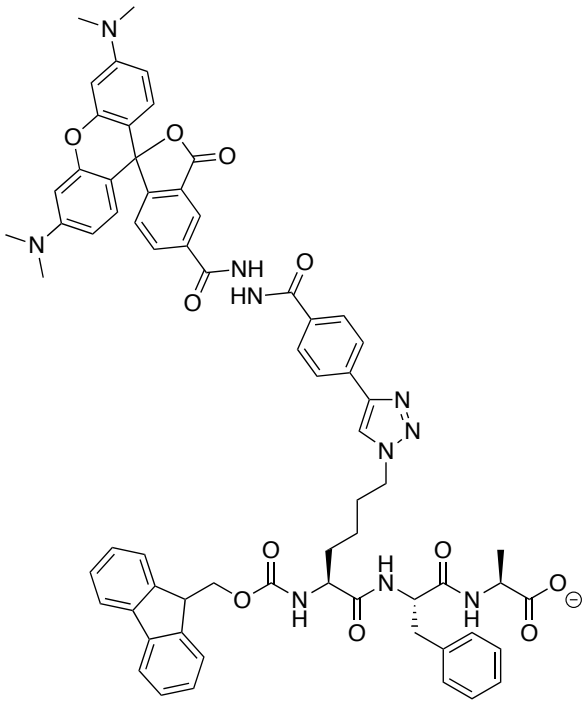
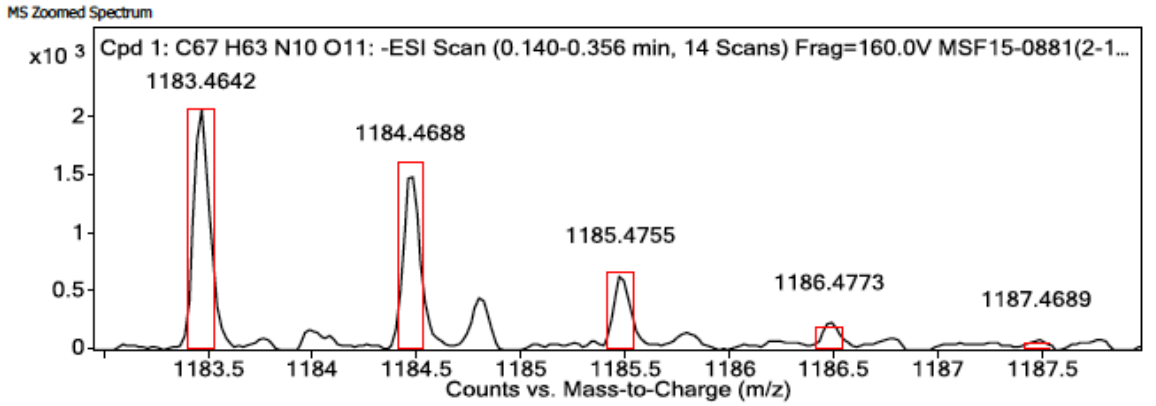


Figure 3.42: (top) Analytical trace for peptide 3.13 at 254 nm; Retention time: 25.78133 min. (bottom) Analytical trace at 550 nm; Retention time: 25.792 min.



(25.9%), 1186.48 (5.3%), 1184.47 (3.7%), 1185.47 (2.7%), 1185.47 (2.3%), 1186.48 (1.6%), 1187.48 (1.0%) m/z: 1183.47 (100.0%), 1184.47 (72.5%), 1185.48

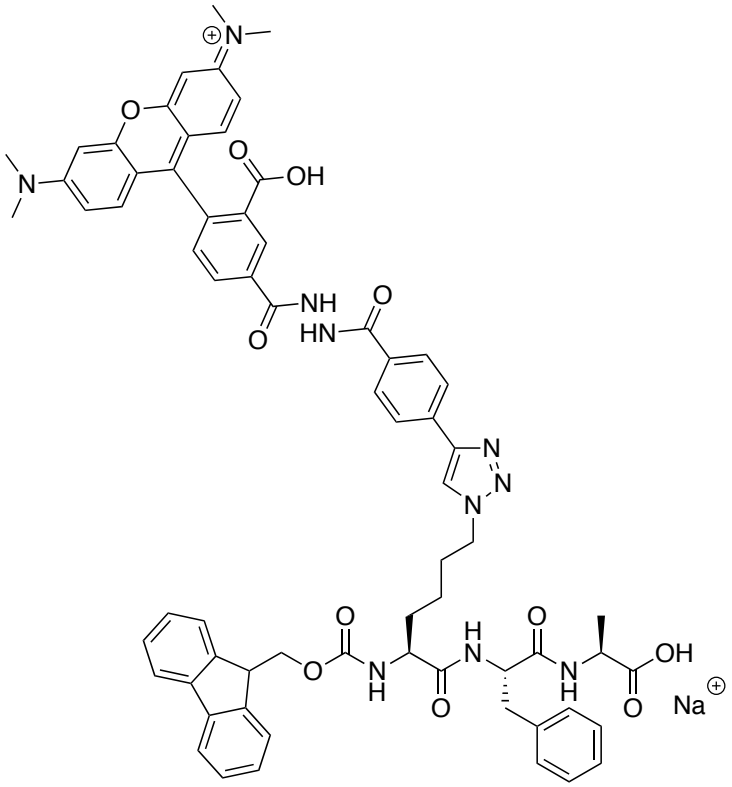
Fmoc-azidolys(hydrazideTMR)-phe-ala



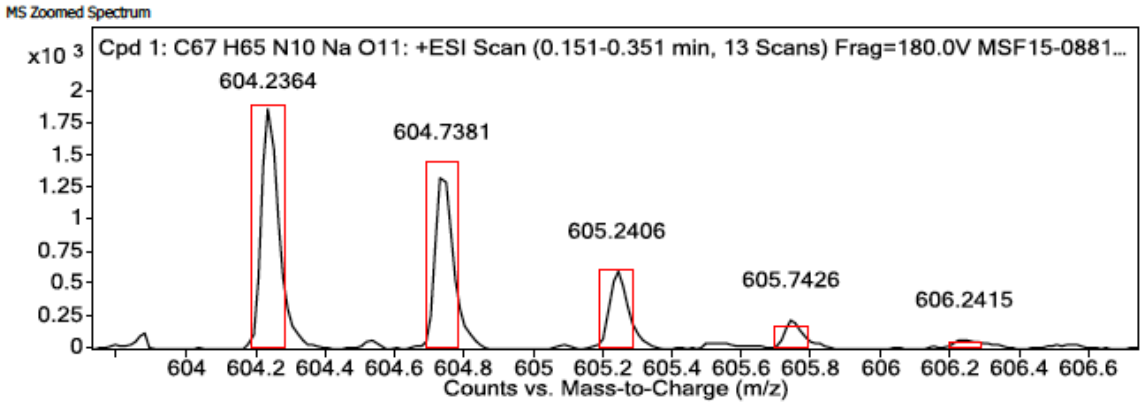
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
520.91010			474028.21			
1183.46420	1183.46830	1	2079.19	C67H63N10O11	M-	3.45
1184.46880	1184.47140	1	1528.58	C67H63N10O11	M-	2.21
1185.47550	1185.47440	1	642.07	C67H63N10O11	M-	-0.97
1186.47730	1186.47720	1	234.02	C67H63N10O11	M-	-0.08
1187.46890	1187.48000	1	88.37	C67H63N10O11	M-	9.37

Figure 3.43: HRMS data for peptide 3.14. Starting amount: 0.015 mmole. Yield: 12%.



m/z: 604.24 (100.0%), 604.74 (72.5%), 605.24 (25.9%), 605.74 (5.3%), 604.73 (3.7%), 605.24 (2.7%), 605.24 (2.3%), 605.74 (1.6%), 606.24 (1.0%)



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
360.32360			117638.59			
604.23640	604.23610	2	1870.13	C67H65N10NaO11	M+2	-0.51
604.73810	604.73760	2	1378.73	C67H65N10NaO11	M+2	-0.81
605.24060	605.23910	2	605.62	C67H65N10NaO11	M+2	-2.53
605.74260	605.74050	2	225.61	C67H65N10NaO11	M+2	-3.4
606.24150	606.24190	2	69.53	C67H65N10NaO11	M+2	0.56

Figure 3.44: HRMS data for peptide 3.14.

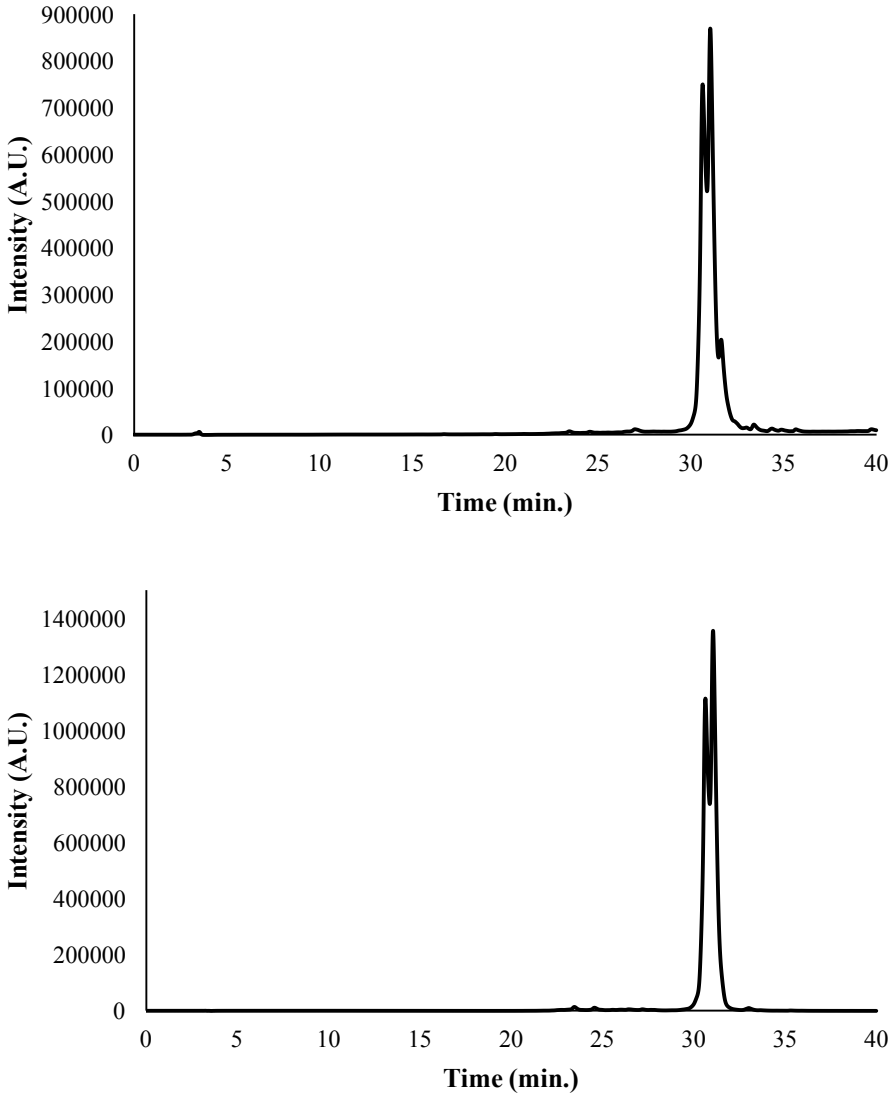
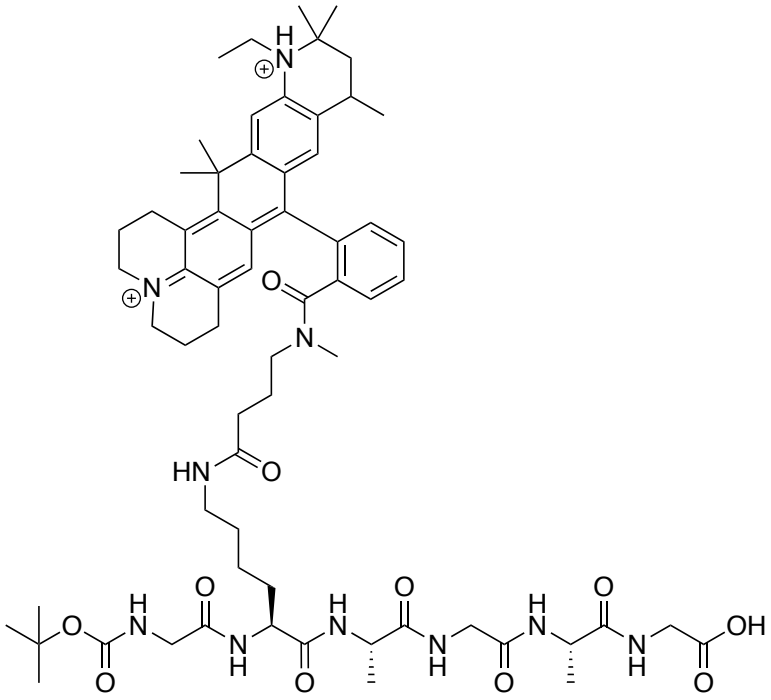
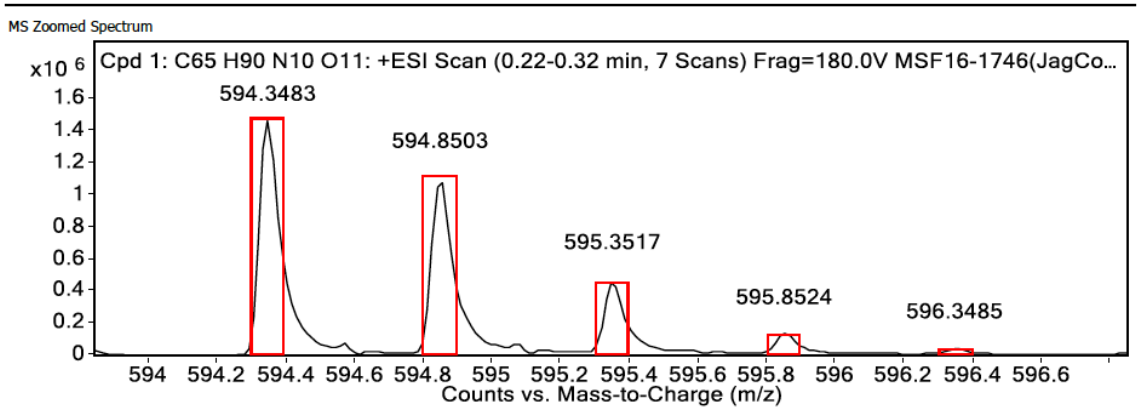


Figure 3.45: (top) Analytical trace for peptide 3.14 at 254 nm; Retention time: 30.61333 min., 31.02933 min. (bottom) Analytical trace at 550 nm; Retention time: 30.61333 min., 31.04 min.



m/z: 594.35 (100.0%), 594.85 (70.3%), 595.35 (24.3%), 595.85 (5.5%), 594.85 (3.7%), 595.35 (2.6%), 595.35 (2.3%), 595.85 (1.6%), 594.85 (1.1%)

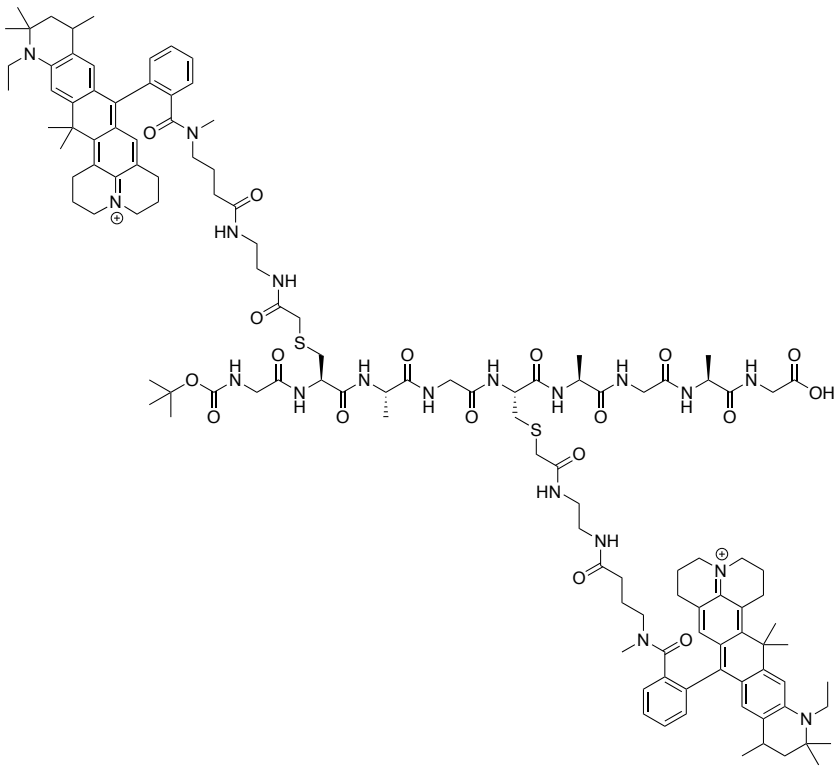
boc-gly-lys(Atto647N)-ala-gly-ala-gly



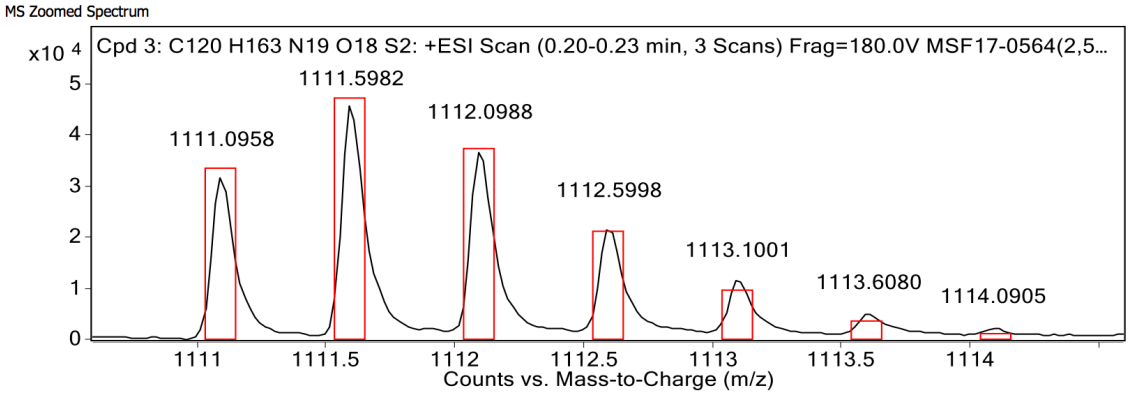
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
594.34830	594.34680	2	1466213.59	C65H90N10O11	(M+2H)+2	-2.49
594.85030	594.84840	2	1098314.15	C65H90N10O11	(M+2H)+2	-3.26
595.35170	595.34980	2	464993.48	C65H90N10O11	(M+2H)+2	-3.23
595.85240	595.85120	2	121367.26	C65H90N10O11	(M+2H)+2	-1.85
596.34850	596.35260	2	32307.59	C65H90N10O11	(M+2H)+2	6.98
596.83850	596.85400	2	9897	C65H90N10O11	(M+2H)+2	26.02
605.33910			6987045.11			

Figure 3.46: HRMS data for peptide 3.15. Starting amount: 0.0014 mmole. Yield: 54%.



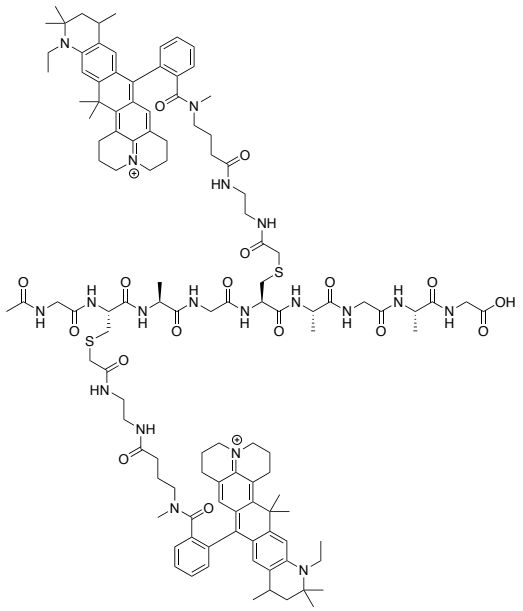
m/z: 1111.59 (100.0%), 1111.09 (77.0%), 1112.10 (32.9%), 1112.10 (31.4%), 1112.60 (22.1%), 1112.59 (9.0%), 1113.10 (8.1%), 1112.09 (7.0%), 1112.09 (7.0%), 1111.59 (5.4%), 1112.60 (5.3%), 1112.59 (4.5%), 1113.09 (3.0%), 1113.09 (2.8%), 1112.60 (2.3%), 1113.60 (2.1%), 1113.60 (2.0%), 1112.10 (1.9%), 1112.09 (1.7%), 1112.09 (1.6%), 1111.60 (1.4%), 1112.60 (1.4%), 1113.10 (1.3%), 1111.59 (1.2%), 1113.10 (1.2%), 1113.10 (1.2%), 1112.09 (1.1%)



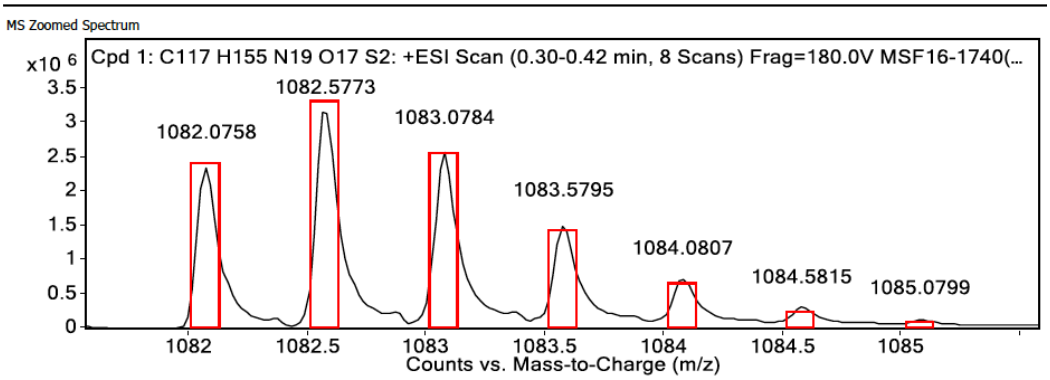
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
304.26250			289572.25			
1111.09580	1111.09270	2	31990.93	C120H163N19O18S2	M+2	-2.79
1111.59820	1111.59420	2	46307.9	C120H163N19O18S2	M+2	-3.61
1112.09880	1112.09530	2	37326.38	C120H163N19O18S2	M+2	-3.17
1112.59980	1112.59620	2	22017.88	C120H163N19O18S2	M+2	-3.23
1113.10010	1113.09700	2	10321.63	C120H163N19O18S2	M+2	-2.84
1113.60800	1113.59770	2	4021.68	C120H163N19O18S2	M+2	-9.27
1114.09050	1114.09850	2	1301.98	C120H163N19O18S2	M+2	7.12
1114.66150	1114.59920	2	660.58	C120H163N19O18S2	M+2	-55.91

Figure 3.47: HRMS data for peptide 3.16.



m/z: 1082.57 (100.0%), 1082.07 (79.0%), 1083.08 (62.7%), 1083.58 (16.8%), 1083.58 (9.2%), 1083.57 (9.0%), 1083.07 (7.1%), 1083.07 (7.0%), 1084.07 (5.7%), 1082.57 (ac-gly-cys(Atto647N)-ala-gly-cys(Atto647N)-ala-gly-ala-gly



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
1082.07580	1082.07180	2	2347442.9	C117H155N19O17S2	(M+2H)+2	-3.75
1082.57730	1082.57330	2	3246090.92	C117H155N19O17S2	(M+2H)+2	-3.72
1083.07840	1083.07430	2	2561994.98	C117H155N19O17S2	(M+2H)+2	-3.73
1083.57950	1083.57520	2	1503524.17	C117H155N19O17S2	(M+2H)+2	-3.96
1084.08070	1084.07590	2	652302.18	C117H155N19O17S2	(M+2H)+2	-4.42
1084.58150	1084.57670	2	231696.57	C117H155N19O17S2	(M+2H)+2	-4.47
1085.07990	1085.07740	2	72628.91	C117H155N19O17S2	(M+2H)+2	-2.26
1085.58940	1085.57820	2	20151.3	C117H155N19O17S2	(M+2H)+2	-10.4
1086.10900	1086.07890	2	6393.48	C117H155N19O17S2	(M+2H)+2	-27.69

Figure 3.48: HRMS data for peptide 3.17. Starting amount: 9.2 x 10^{-5} mmole. Yield: 63%.

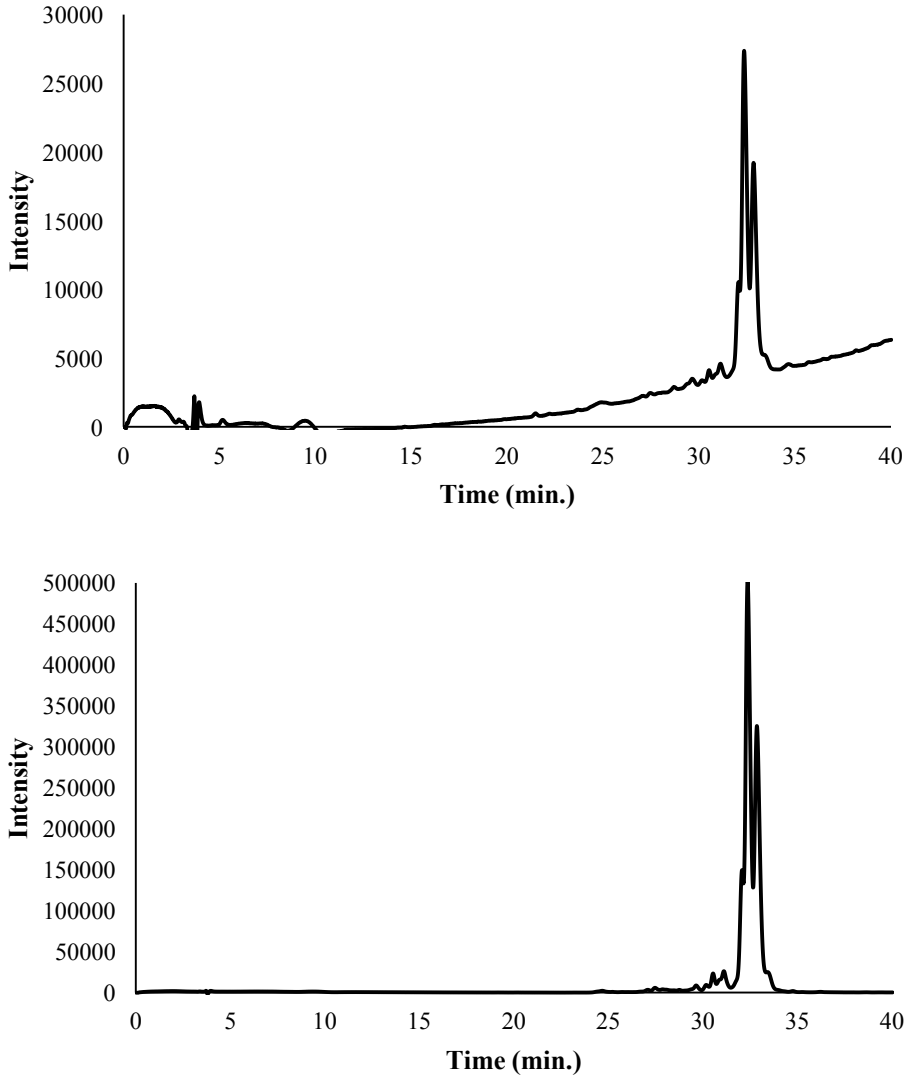
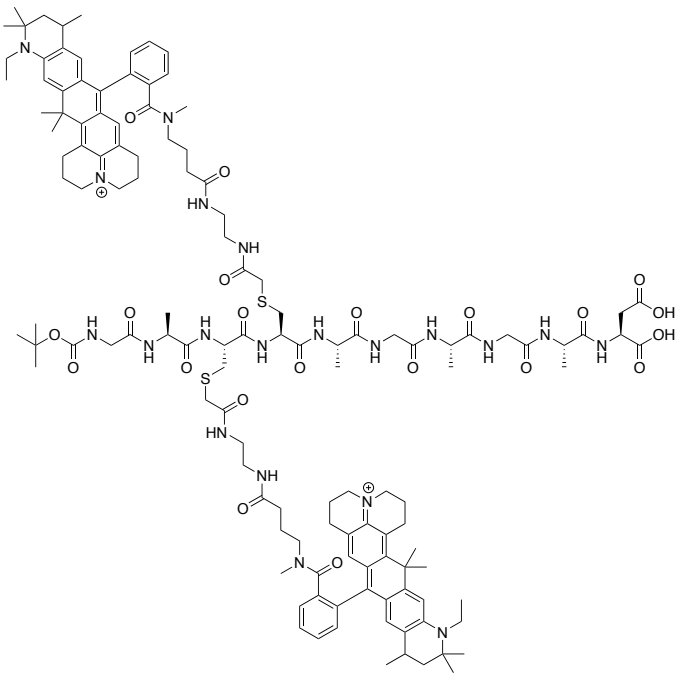
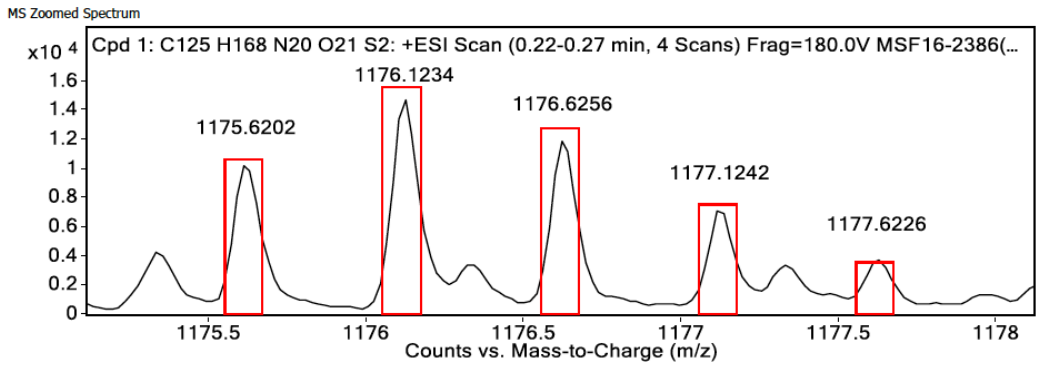


Figure 3.49: (top) Analytical trace for peptide 3.17 at 254 nm; Retention times: 32.0933 min., 32.24533 min. (bottom) Analytical trace at 640 nm; Retention times: 32.32 min., 32.46933 min.



m/z: 1176.12 (100.0%), 1175.61 (74.0%), 1176.62 (67.1%), 1177.12 (21.1%), 1177.62 (9.8%), 1177.11 (9.0%), 1177.12 (8.6%), 1176.61 (7.4%), 1176.61 (6.7%), 1177.62 (6.1%), 1176.11 (5.5%), 1177.12 (5.0%), 1177.12 (4.3%), 1176.62 (3.2%), 1177.62 (2.9%), 1178.12 (2.0%), 1176.62 (2.0%), 1178.12 (1.9%), 1177.62 (1.7%), 1176.62

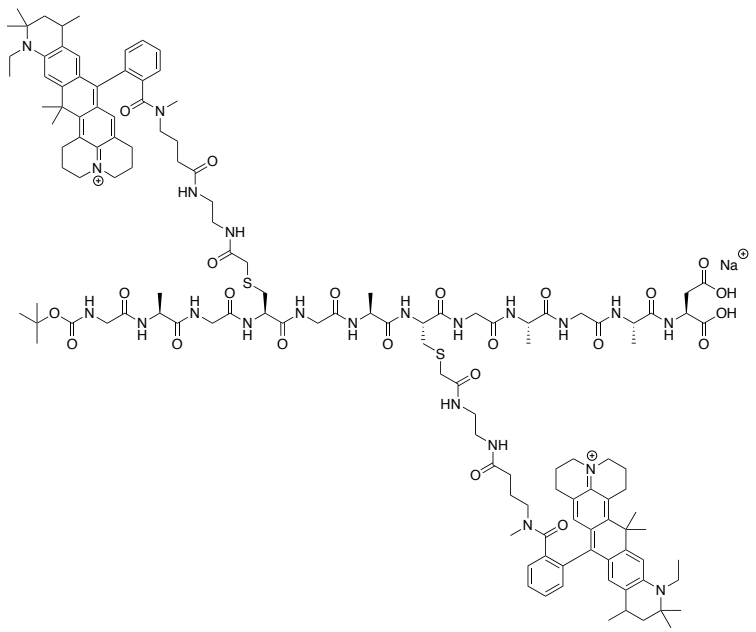
boc-gly-ala-cys(Atto647N)-cys(Atto647N)-ala-gly-ala-gly-ala-asp



MS Spectrum Peak List

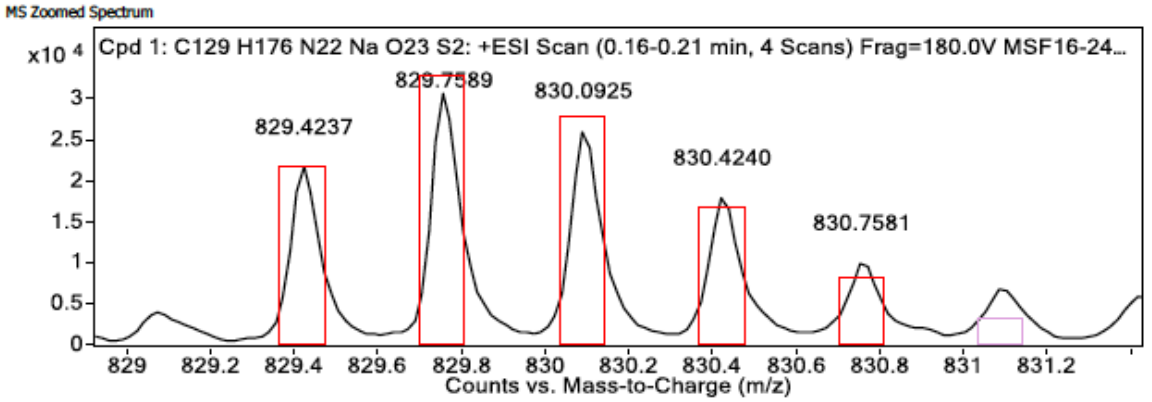
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
525.24910			2789801.05			
1175.62020	1175.61400	2	10429.87	C125H168N20O21S2	(M+2H)+2	-5.25
1176.12340	1176.11550	2	14801.95	C125H168N20O21S2	(M+2H)+2	-6.74
1176.62560	1176.61660	2	12043.79	C125H168N20O21S2	(M+2H)+2	-7.62
1177.12420	1177.11750	2	7317.19	C125H168N20O21S2	(M+2H)+2	-5.66
1177.62260	1177.61830	2	3794.33	C125H168N20O21S2	(M+2H)+2	-3.62
1178.12200	1178.11910	2	2040.25	C125H168N20O21S2	(M+2H)+2	-2.45
1178.60470	1178.61990	2	1228.42	C125H168N20O21S2	(M+2H)+2	12.93

Figure 3.50: HRMS data for peptide 3.18. Starting amount: 6.6 x 10^{-4} mmole. Yield: 44%.



m/z: 829.75 (100.0%), 829.42 (71.7%), 830.09 (69.2%), 830.42 (31.7%), 830.42 (9.0%), 830.09 (8.1%), 830.09 (6.5%), 830.75 (6.3%), 830.76 (6.0%), 829.75 (5.8%), 830.76

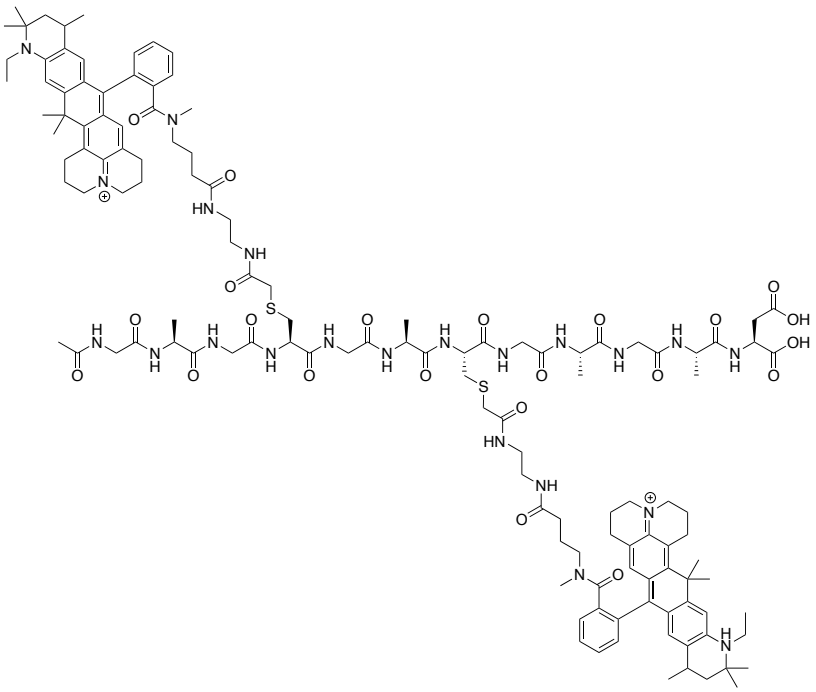
boc-gly-ala-gly-cys(Atto647N)-gly-ala-cys(Atto647N)-gly-ala-gly-ala-asp



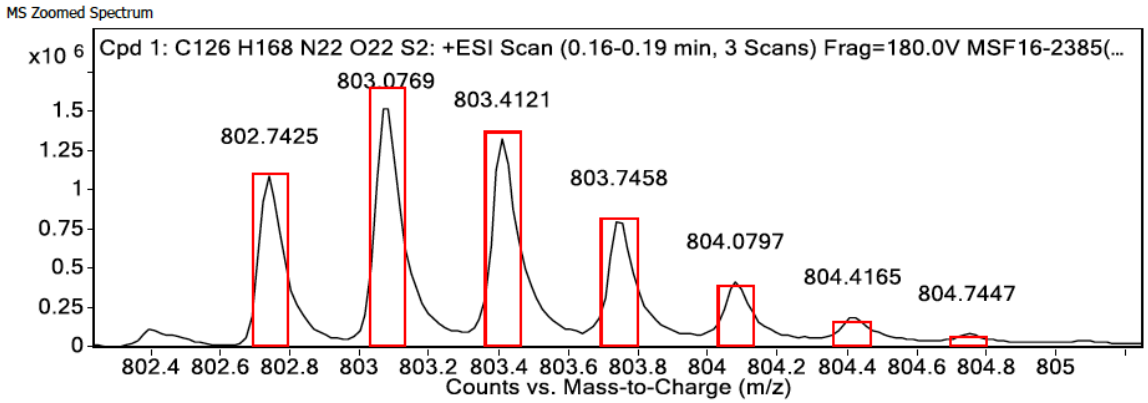
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
130.15920			128331.74			
829.42370	829.42000	3	21899.5	C129H176N22NaO23S2	M+3	-4.46
829.75890	829.75440	3	30870.21	C129H176N22NaO23S2	M+3	-5.47
830.09250	830.08850	3	26307.4	C129H176N22NaO23S2	M+3	-4.82
830.42400	830.42240	3	18137.9	C129H176N22NaO23S2	M+3	-1.95
830.75810	830.75630	3	10193.11	C129H176N22NaO23S2	M+3	-2.11

Figure 3.51: HRMS data for peptide 3.19. Starting material: 7.5 x 10^{-4}. Yield: 35%.



m/z: 803.08 (100.0%), 802.75 (73.4%), 803.41 (67.6%), 803.75 (23.1%), 804.08 (9.1%), 803.75 (9.0%), 803.41 (8.1%), 803.75 (7.1%), 803.41 (6.6%), 803.08 (6.0%), 803.75 (5.5%), 803.75 (4.5%), 803.41 (3.3%), 804.08 (3.3%), 804.08 (3.1%), 804.08 (2.8%), ac-gly-ala-gly-cys(Atto647N)-gly-ala-cys(Atto647N)-gly-ala-gly-ala-asp



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
802.74250	802.74540	3	1093086.46	C126H168N22O22S2	(M+3H)+3	3.68
803.07690	803.07980	3	1558838.37	C126H168N22O22S2	(M+3H)+3	3.65
803.41210	803.41380	3	1327036.88	C126H168N22O22S2	(M+3H)+3	2.16
803.74580	803.74780	3	827149.93	C126H168N22O22S2	(M+3H)+3	2.46
804.07970	804.08170	3	422377.63	C126H168N22O22S2	(M+3H)+3	2.47
804.41650	804.41550	3	196485.89	C126H168N22O22S2	(M+3H)+3	-1.26
804.74470	804.74940	3	89842.28	C126H168N22O22S2	(M+3H)+3	5.77

Figure 3.52: HRMS data for peptide 3.20. Starting amount: 1.3 x 10^{-3}. Yield: 71%.

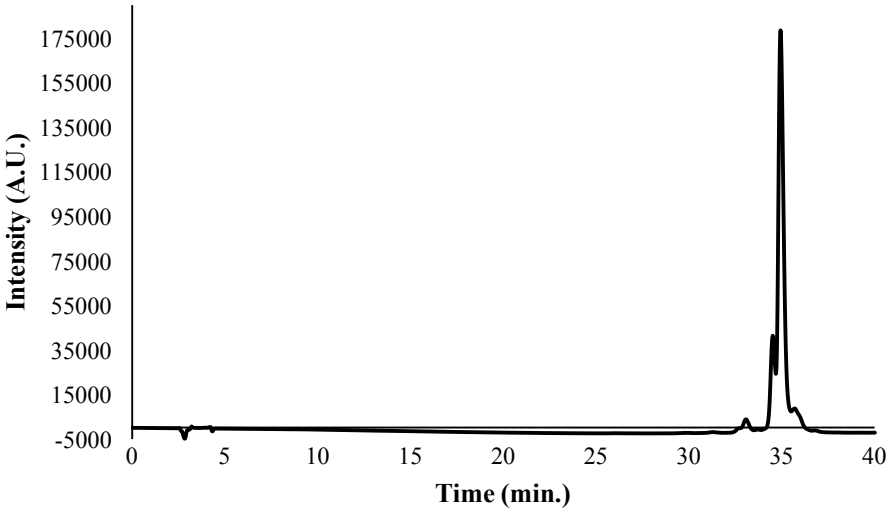
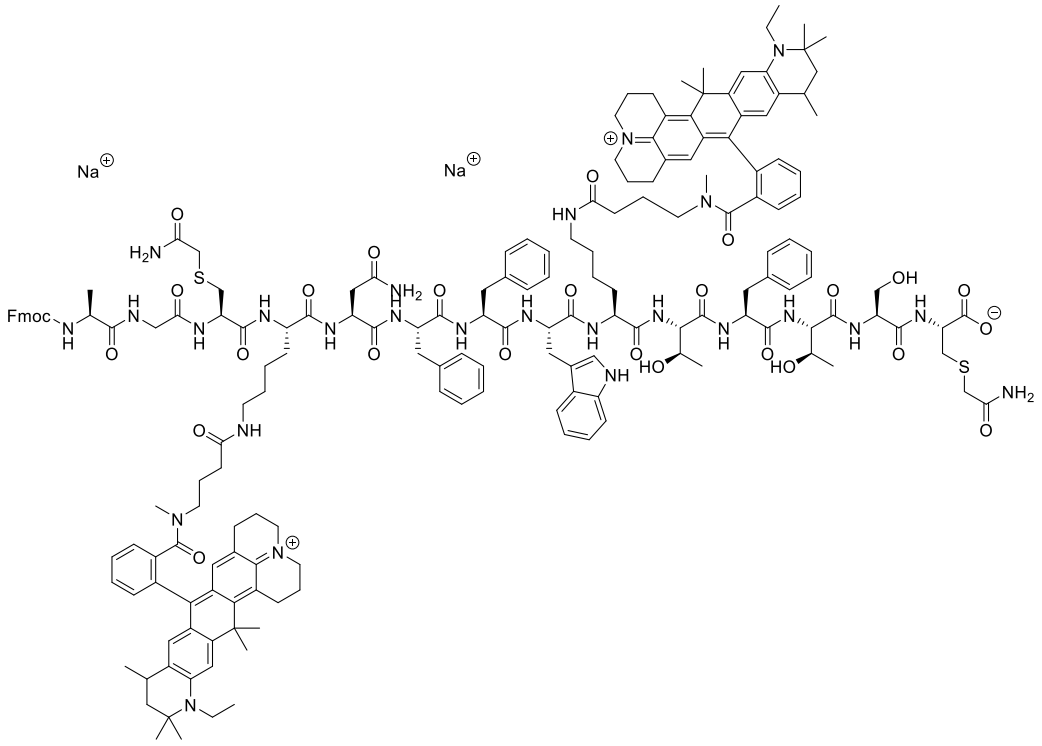


Figure 3.53: Analytical trace for peptide 3.20 at 640 nm; Retention time: 34.912 min.



m/z: 1092.53 (100.0%), 1092.87 (96.3%), 1093.20 (61.4%), 1092.20 (51.7%), 1093.54 (19.6%), 1092.86 (9.6%), 1093.54 (9.6%), 1093.20 (9.2%), 1093.20 (9.0%), 1093.53 (8.7%), 1093.87 (7.9%), 1093.20 (5.5%), 1093.53 (5.3%), 1092.53 (5.0%), 1092.86 (4.7%), 1093.53 (3.3%), 1093.87 (3.1%), 1093.87 (3.1%), 1092.87 (2.9%), 1093.53 (2.6%), 1092.87 (2.5%), 1093.20 (2.4%), 1093.87 (2.4%), 1094.20 (2.3%), 1093.87 (1.9%), 1093.87 (1.9%), 1094.20 (1.8%), 1092.87 (1.6%), 1093.20 (1.5%), 1093.87 (1.5%), 1092.53 (1.3%), 1094.20 (1.1%), 1094.20 (1.1%), 1092.87 (1.0%)

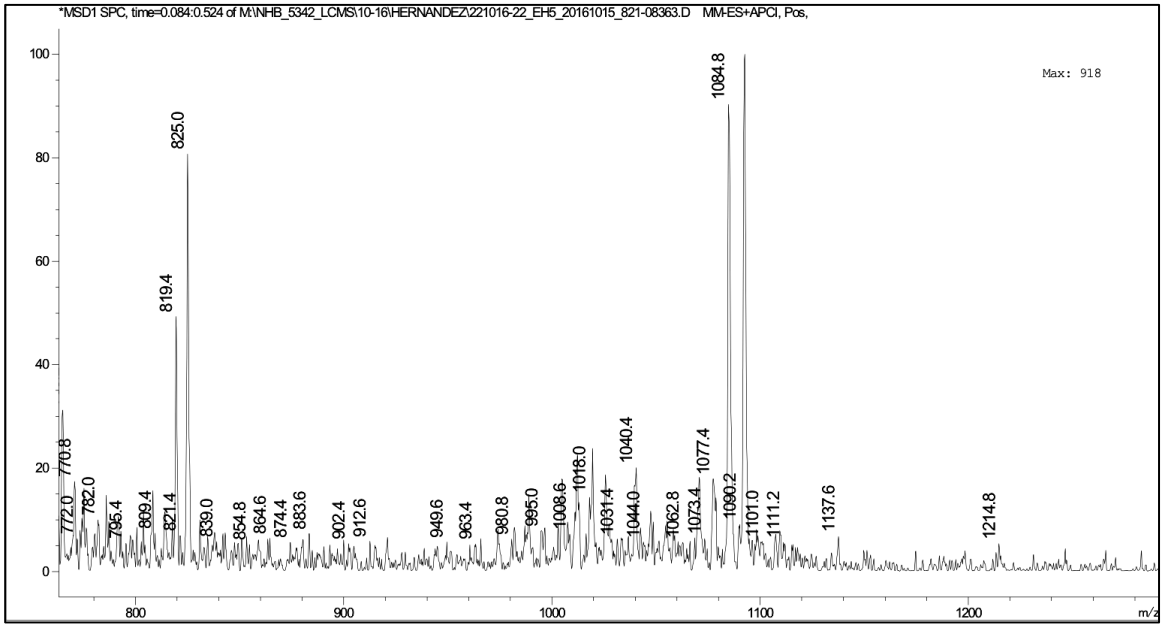
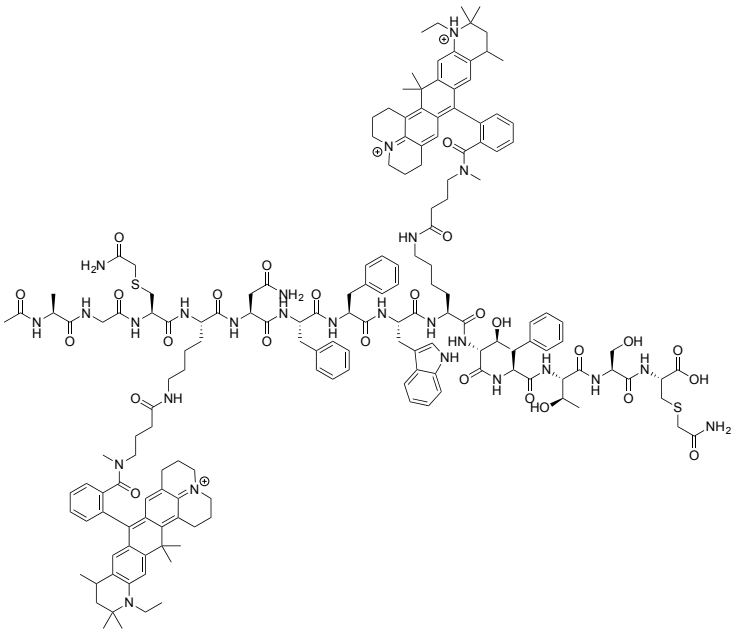


Figure 3.54: LRMS data for peptide 3.21. Starting amount: 4.2 x 10^{-4} mmole. Yield: 65%.



m/z: 1017.86 (100.0%), 1018.19 (65.4%), 1017.52 (55.7%), 1018.53 (46.2%), 1018.19 (23.8%), 1018.86 (16.7%), 1018.19 (9.6%), 1018.52 (9.0%), 1018.86 (6.6%), 1018.53 (6.5%), 1018.53 (6.3%), 1018.86 (5.9%), 1017.86 (5.3%), 1018.53 (5.3%), 1018.19 (5.0%), 1018.86 (4.4%), ac-ala-gly-cys-lys(Atto647N)-asn-phe-phe-trp-lys(Atto647N)-thr-phe-thr-ser-cys

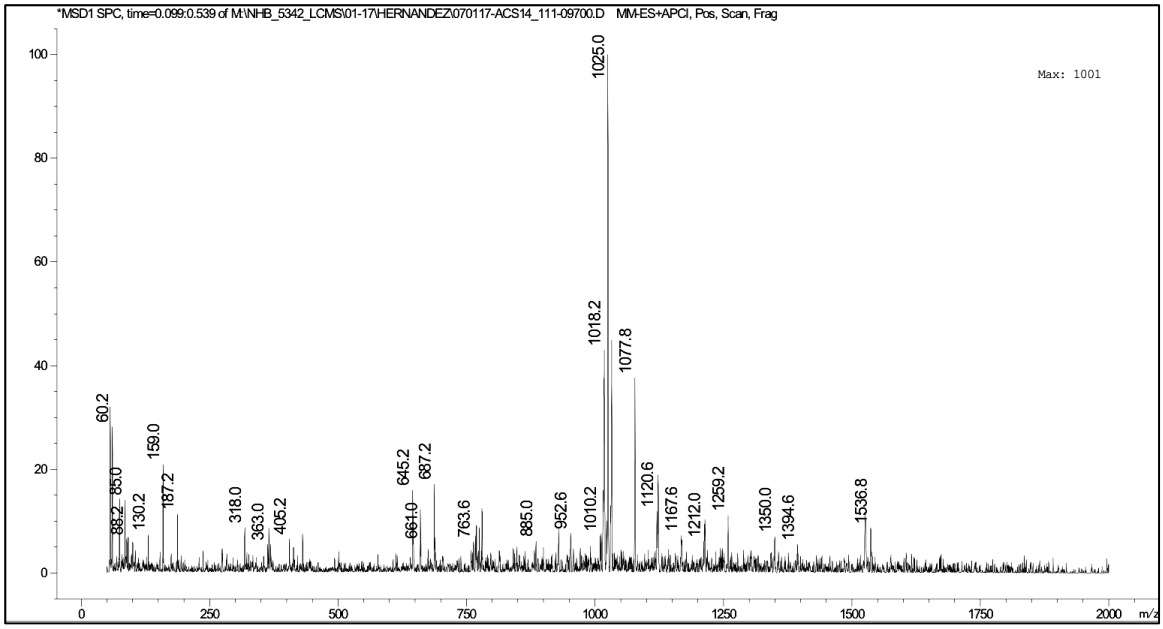
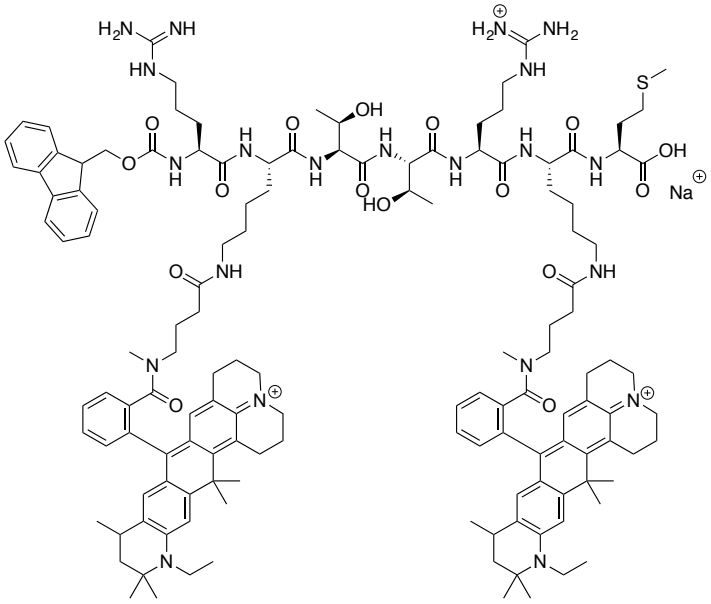


Figure 3.55: LRMS data for peptide 3.22. Starting amount: 9.0 x 10^{-4} mmole. Yield: 83%.



m/z: 605.85 (100.0%), 606.10 (73.0%), 605.60 (68.0%), 606.35 (27.3%), 606.60 (12.7%), 606.35 (7.9%), 606.10 (7.8%), 606.35 (5.7%), 605.84 (5.3%), 606.35 (4.5%), 606.60 (3.3%), 606.35 (3.3%), 606.09 (3.1%), 606.85 (2.5%), 606.60 (2.4%), 606.10 (2.2%), 606.10 (2.1%), 606.60 (1.6%), 606.35 (1.5%), 605.85 (1.4%), 606.85 (1.2%), 606.60 (1.1%), 606.85 (1.1%)

Fmoc-arg-lys(Atto647N)-thr-thr-arg-lys(Atto647N)-met

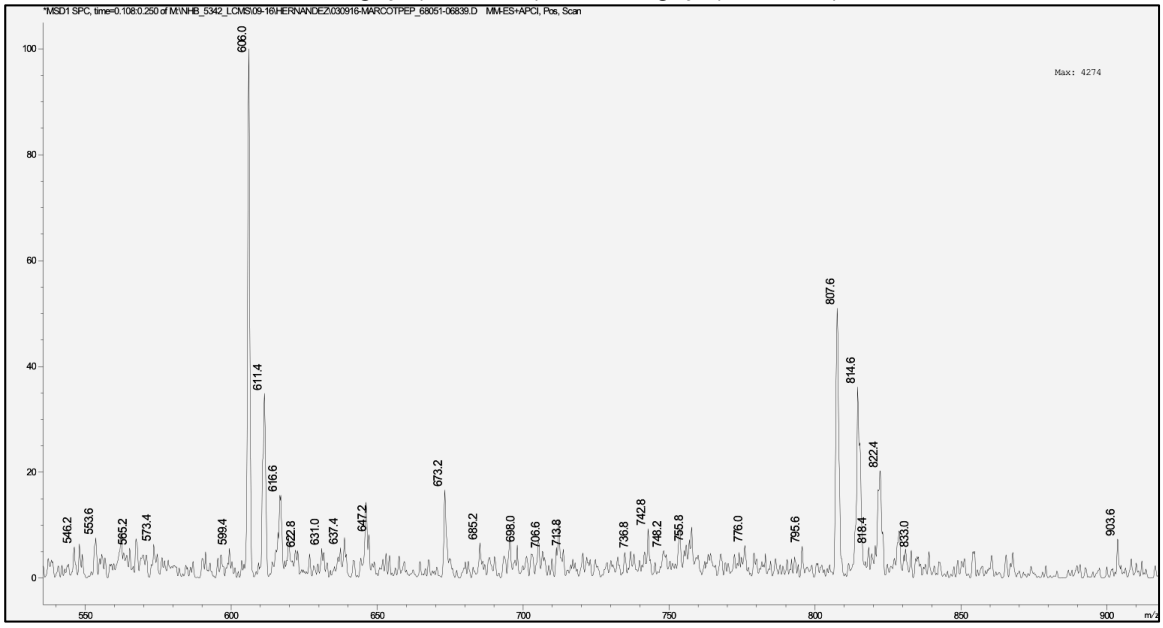
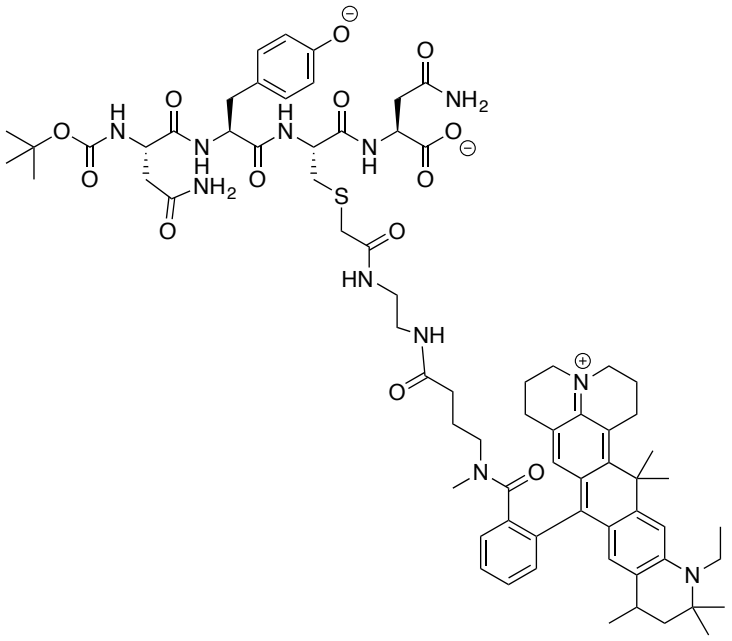
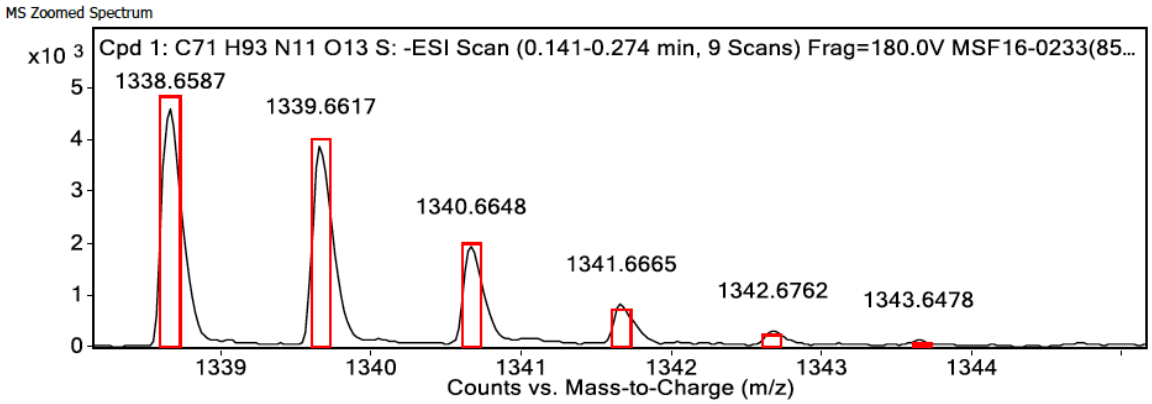


Figure 3.56: LRMS data for peptide 3.23. Starting amount: 5.6 x 10^{-4} mmole. Yield: 33%.



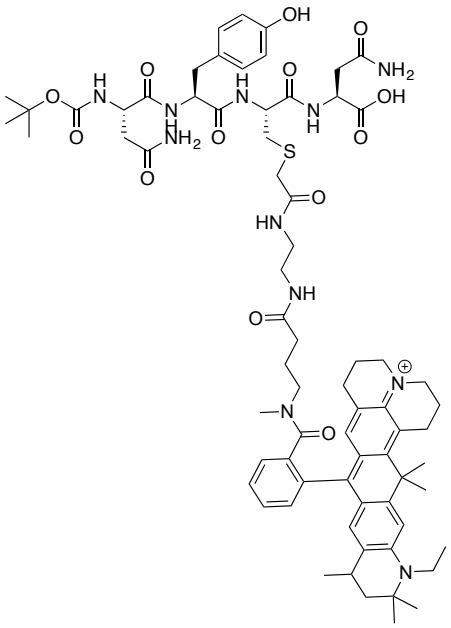
m/z: 1338.66 (100.0%), 1339.66 (76.8%), 1340.67 (29.1%), 1341.67 (6.4%), 1340.66 (4.5%), 1339.66 (4.1%), 1341.66 (3.5%), 1340.66 (2.8%), 1340.66 (2.7%), 1341.67 (2.1%), 1342.67 (1.3%), 1342.66 (1.3%), 1341.66 (1.2%), 1339.67 (1.1%)



MS Spectrum Peak List

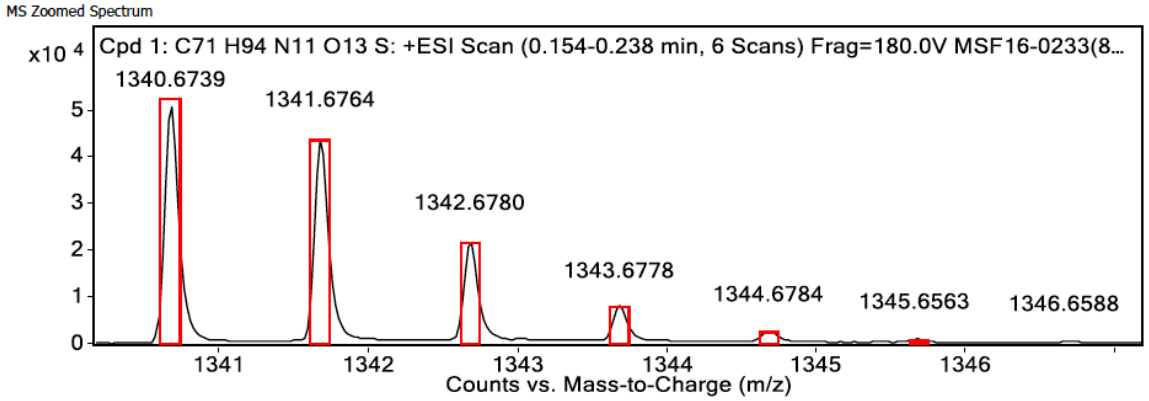
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
980.01910			148978.29			
1338.65870	1338.66020	1	4627.87	C71H93N11O13S	(M-H)-	1.11
1339.66170	1339.66330	1	3903.36	C71H93N11O13S	(M-H)-	1.17
1340.66480	1340.66510	1	1963.95	C71H93N11O13S	(M-H)-	0.18
1341.66650	1341.66660	1	852.66	C71H93N11O13S	(M-H)-	0.1
1342.67620	1342.66820	1	339.71	C71H93N11O13S	(M-H)-	-5.96
1343.64780	1343.66990	1	161.38	C71H93N11O13S	(M-H)-	16.42

Figure 3.57: HRMS data for peptide 3.24. Starting amount: 1.1 x 10^{-3} mmole. Yield: 28%.



m/z: 1340.67 (100.0%), 1341.68 (76.8%), 1342.68 (29.1%), 1343.68 (6.4%), 1342.67 (4.5%), 1341.67 (4.1%), 1343.67 (3.5%), 1342.68 (2.8%), 1342.68 (2.7%), 1343.68 (2.1%), 1344.69 (1.3%), 1344.68 (1.3%), 1343.68 (1.2%), 1341.68 (1.1%)

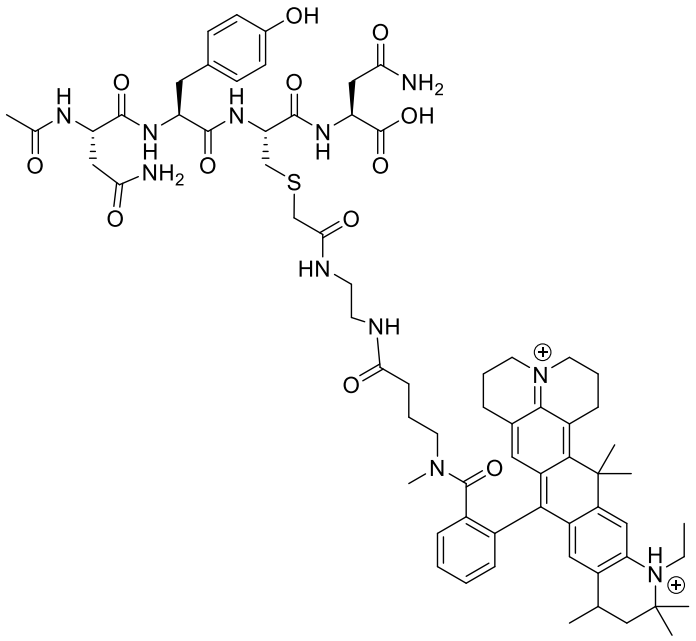
boc-asn-tyr-cys(Atto647N)-asn



MS Spectrum Peak List

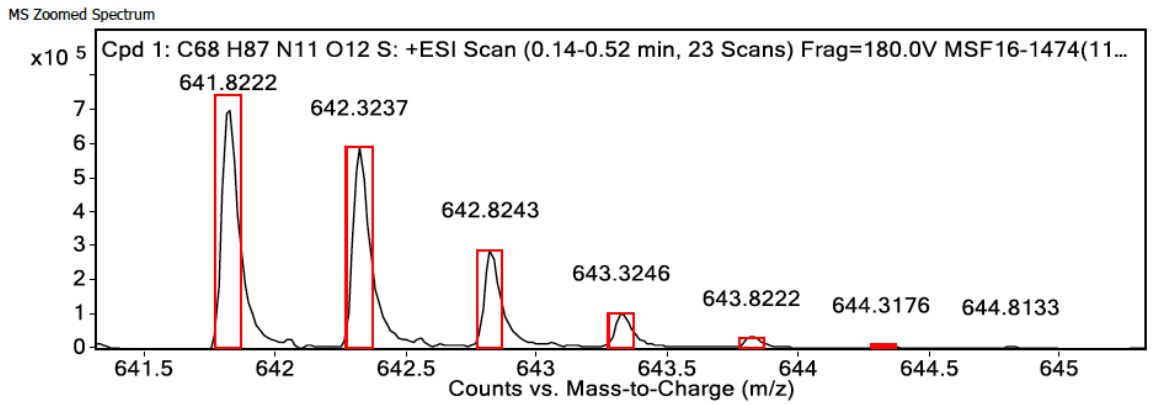
Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
681.83640			1723745.34			
1340.67390	1340.67480	1	51266.14	C71H94N11O13S	M+	0.66
1341.67640	1341.67780	1	44059.6	C71H94N11O13S	M+	1.05
1342.67800	1342.67960	1	22008.63	C71H94N11O13S	M+	1.2
1343.67780	1343.68120	1	7478.75	C71H94N11O13S	M+	2.51
1344.67840	1344.68270	1	2438.06	C71H94N11O13S	M+	3.21
1345.65630	1345.68450	1	854.21	C71H94N11O13S	M+	20.95
1346.65880	1346.68630	1	382.84	C71H94N11O13S	M+	20.43
1347.73000	1347.68830	1	175.95	C71H94N11O13S	M+	-30.93

Figure 3.58: HRMS data for peptide 3.24.



m/z: 641.82 (100.0%), 642.32 (73.5%), 642.82 (26.6%), 643.33 (6.3%), 642.82 (4.5%), 642.32 (3.7%), 643.32 (3.3%), 642.82 (3.0%), 642.82 (2.5%), 643.32 (1.8%), 643.82 (1.2%), 642.32 (1.0%)

ac-asn-tyr-cys(Atto647N)-asn



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
641.82220	641.82010	2	715628.09	C68H87N11O12S	(M+2H)+2	-3.25
642.32370	642.32160	2	591495.8	C68H87N11O12S	(M+2H)+2	-3.22
642.82430	642.82250	2	291317.72	C68H87N11O12S	(M+2H)+2	-2.88
643.32460	643.32320	2	105426.7	C68H87N11O12S	(M+2H)+2	-2.17
643.82220	643.82400	2	31465.78	C68H87N11O12S	(M+2H)+2	2.78
644.31760	644.32480	2	11424	C68H87N11O12S	(M+2H)+2	11.2
644.81330	644.82570	2	4990.68	C68H87N11O12S	(M+2H)+2	19.28
645.30880	645.32670	2	2350.68	C68H87N11O12S	(M+2H)+2	27.73
645.80520	645.82770	2	1690.14	C68H87N11O12S	(M+2H)+2	34.92
652.81290			1017861.44			

Figure 3.59: HRMS data for peptide 3.25. Starting amount: 2.3 x 10^{-3} mmole. Yield: 54%.

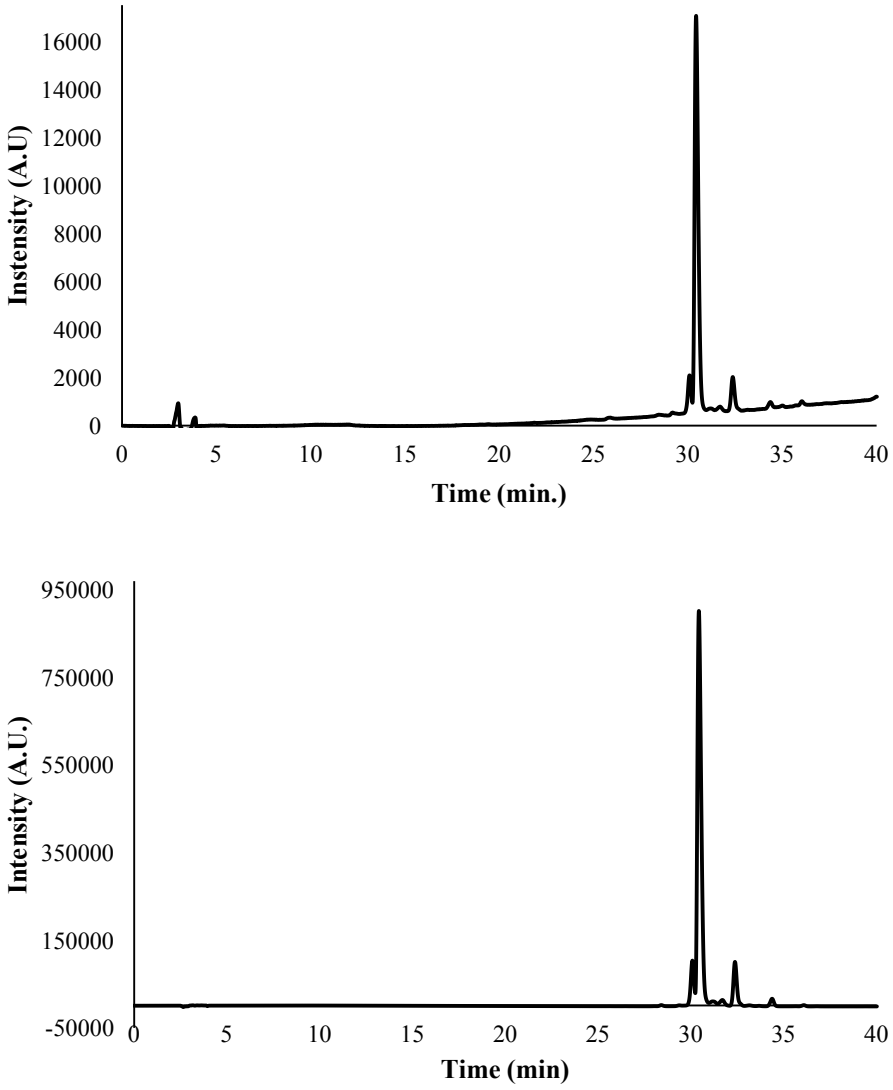
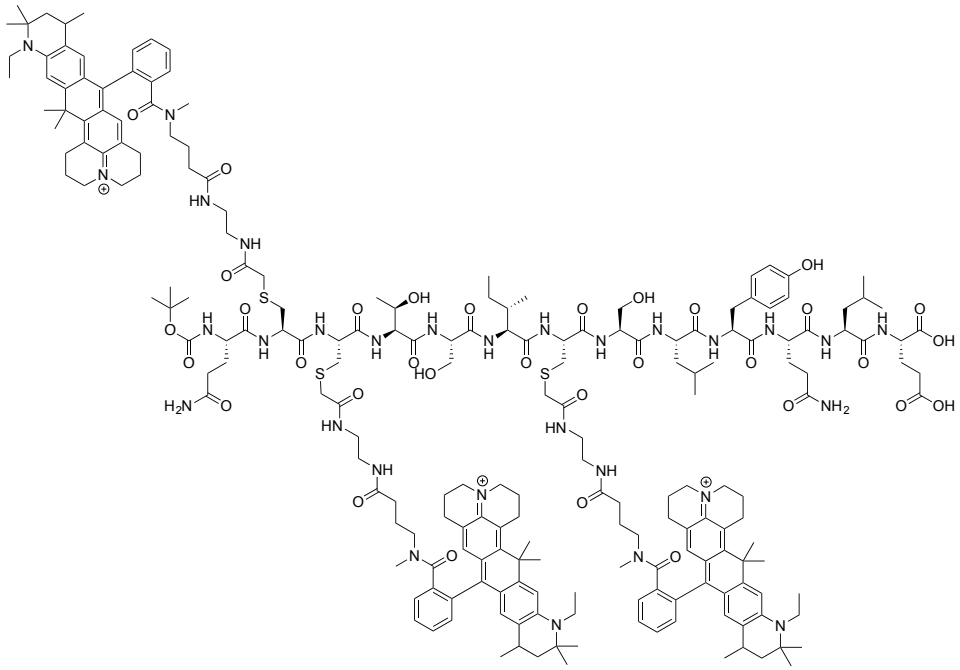
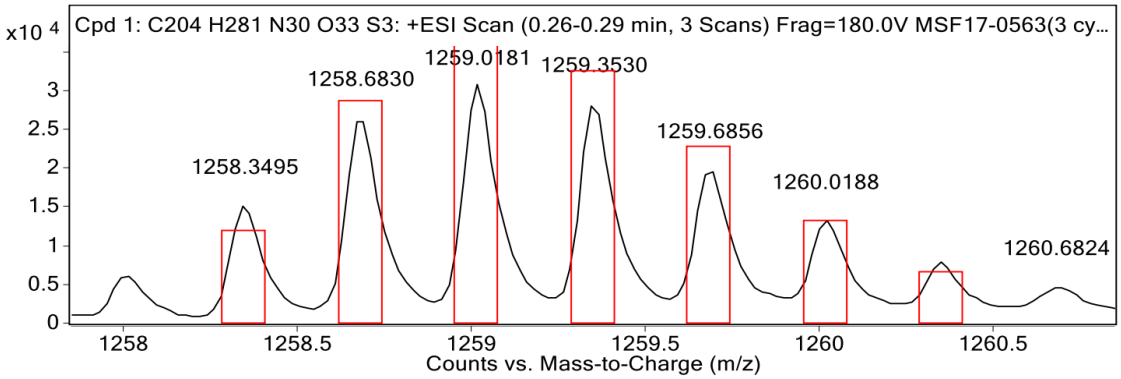


Figure 3.60: (top) Analytical trace for peptide 3.25 at 254 nm; Retention time: 30.41067 min. (bottom) Analytical trace at 640 nm; Retention time: 30.42133 min.



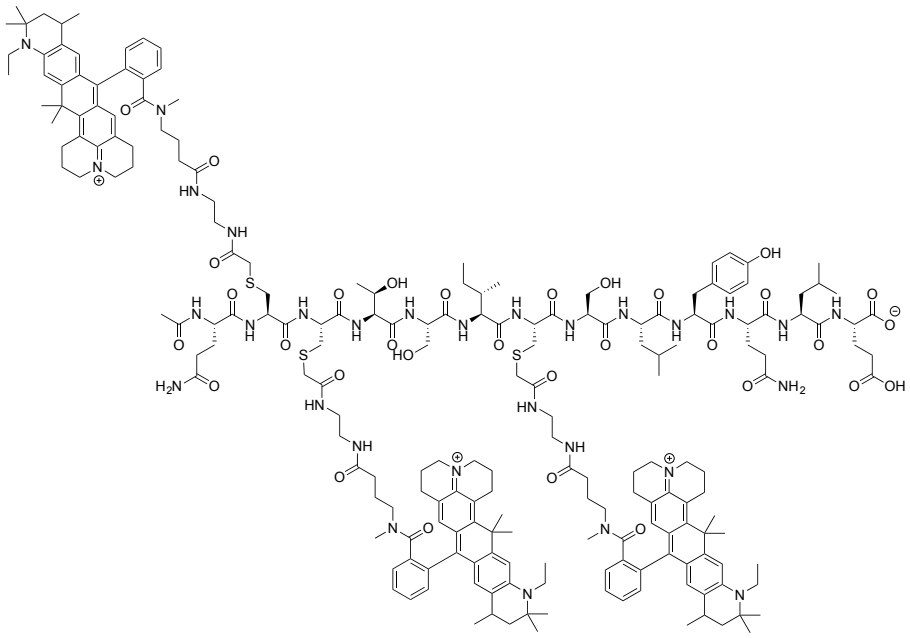
boc-gln-cys(Atto647N)-cys(Atto647N)-thr-ser-ile-cys(Atto647N)-ser-leu-tyr-gln-leu-asp



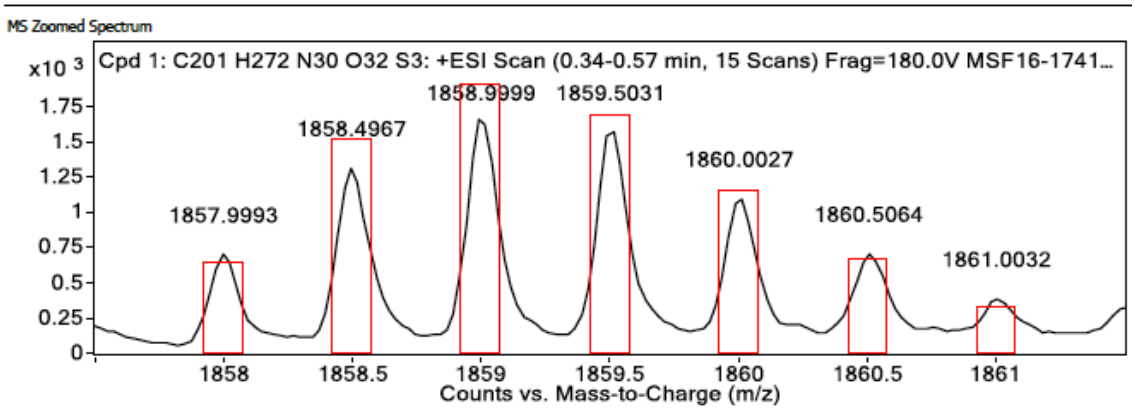
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
994.40910			1474424.82			
1258.34950	1258.34590	3	15371.49	C204H281N30O33S3	M+3	-2.83
1258.68300	1258.68030	3	26715.04	C204H281N30O33S3	M+3	-2.12
1259.01810	1259.01450	3	30886.3	C204H281N30O33S3	M+3	-2.91
1259.35300	1259.34860	3	28517.18	C204H281N30O33S3	M+3	-3.52
1259.68560	1259.68260	3	19971.11	C204H281N30O33S3	M+3	-2.43
1260.01880	1260.01650	3	13415.46	C204H281N30O33S3	M+3	-1.77
1260.34750	1260.35050	3	8132.2	C204H281N30O33S3	M+3	2.32
1260.68240	1260.68440	3	4887.35	C204H281N30O33S3	M+3	1.54
1261.00020	1261.01830	3	3082.14	C204H281N30O33S3	M+3	14.33

Figure 3.61: HRMS data for peptide 3.26. Starting amount: 4.0 x 10^{-4} mmole. Yield: 59%.



m/z: 1859.00 (100.0%), 1858.50 (92.5%), 1859.50 (71.7%), 1857.99 (42.5%), 1860.00 (25.0%),1860.00 (13.6%), 1860.00 (12.9%), 1859.49 (12.5%), 1859.50 (11.1%), 1858.99
ac-gln-cys(Atto647N)-cys(Atto647N)-cys-thr-ser-ile-cys(Atto647N)-ser-leu-tyr-gln-leu-asp



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
930.00840			772761.44			
1857.99930	1857.99430	2	711.52	C201H272N30O32S3	(M+2H)+2	-2.66
1858.49670	1858.49590	2	1323.5	C201H272N30O32S3	(M+2H)+2	-0.45
1858.99990	1858.99710	2	1691.52	C201H272N30O32S3	(M+2H)+2	-1.49
1859.50310	1859.49820	2	1599.17	C201H272N30O32S3	(M+2H)+2	-2.63
1860.00270	1859.99920	2	1107.05	C201H272N30O32S3	(M+2H)+2	-1.83
1860.50640	1860.50020	2	717.21	C201H272N30O32S3	(M+2H)+2	-3.33
1861.00320	1861.00110	2	397.98	C201H272N30O32S3	(M+2H)+2	-1.15
1861.49210	1861.50190	2	337.92	C201H272N30O32S3	(M+2H)+2	5.3
1861.99960	1862.00270	2	225.03	C201H272N30O32S3	(M+2H)+2	1.7

Figure 3.62: HRMS data for peptide 3.27. Starting amount: 4.1 x 10^{-4} mmole. Yield: 79%.

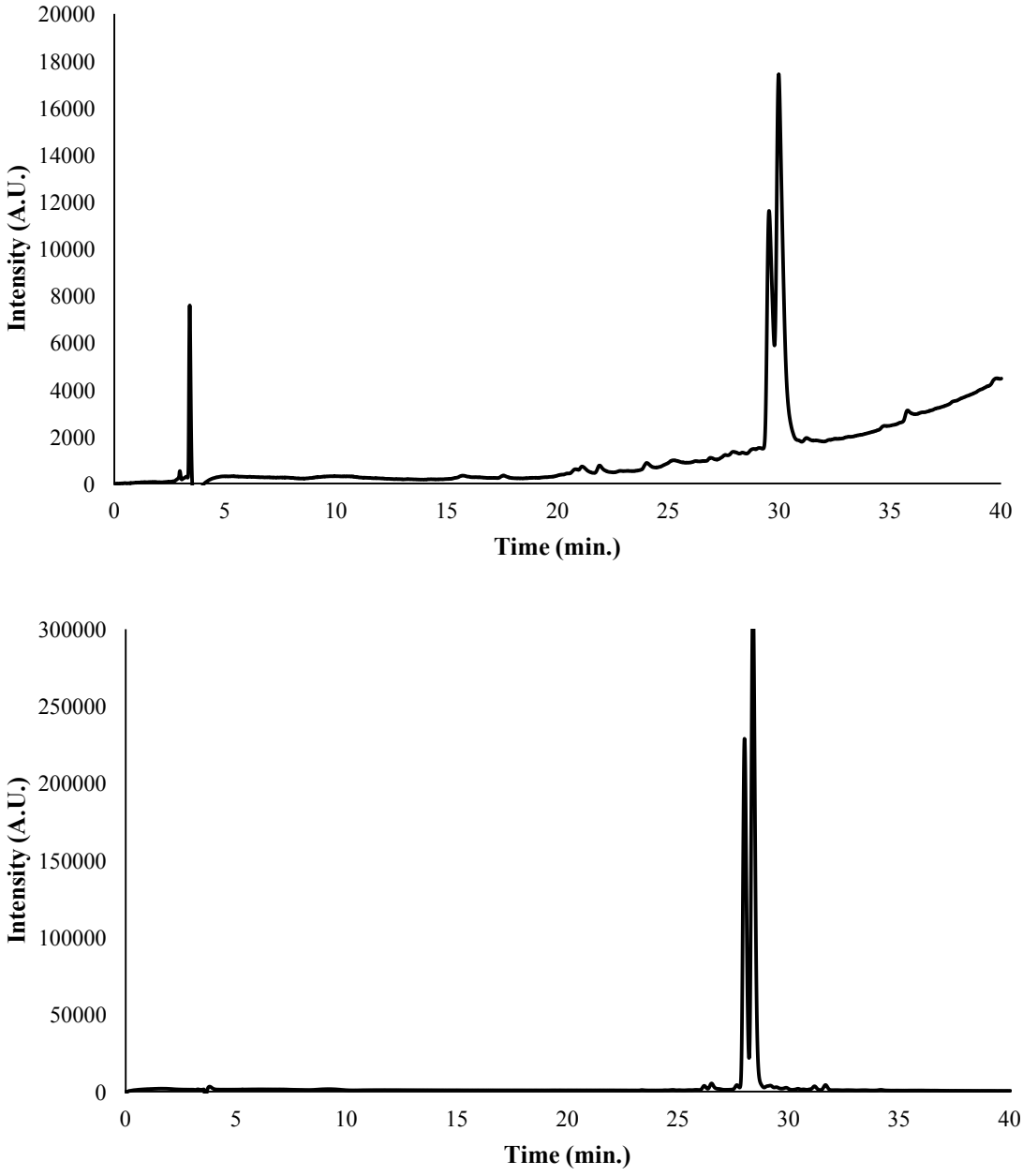
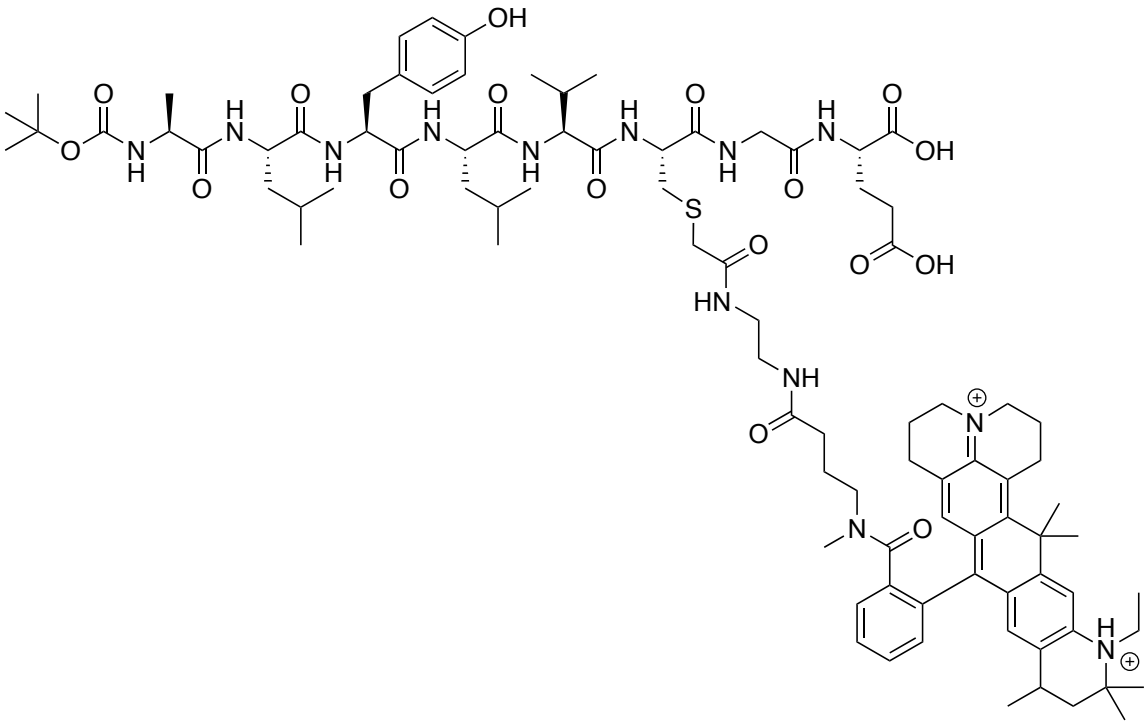
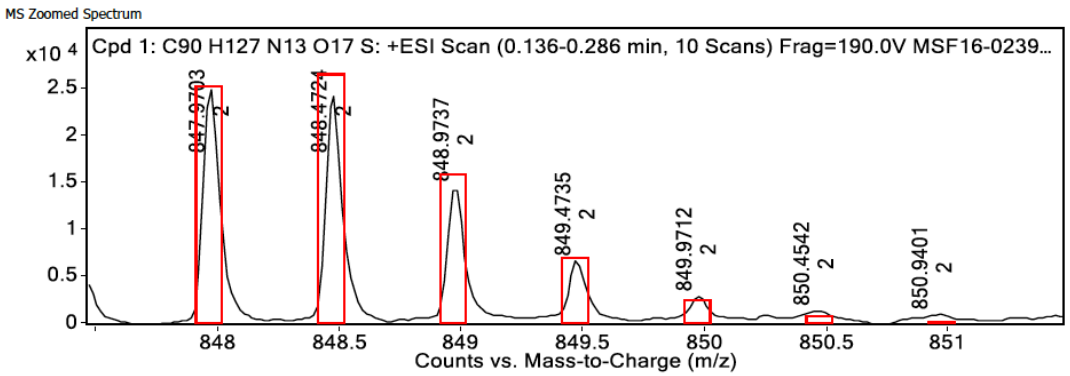


Figure 3.63: (top) Analytical trace for peptide 3.27 at 254 nm; Retention times: 29.504 min., 29.952 min. (bottom) Analytical trace at 640 nm; Retention times: 27.97867 min., 28.34133 min.



m/z: 847.97 (100.0%), 848.47 (97.3%), 848.97 (42.8%), 849.47 (11.7%), 848.47 (4.8%), 848.97 (4.7%), 848.96 (4.5%), 849.47 (4.4%), 848.97 (4.1%), 848.97 (3.5%), 849.47

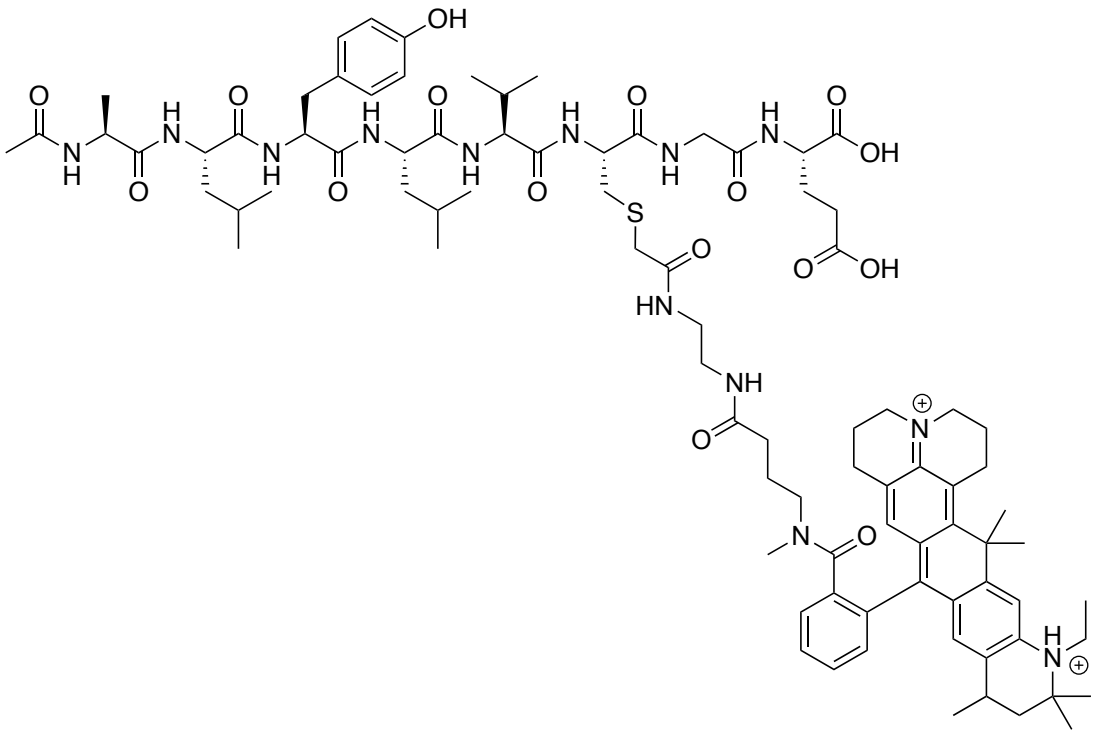
boc-ala-leu-tyr-leu-val-cys(Atto647N)-gly-glu



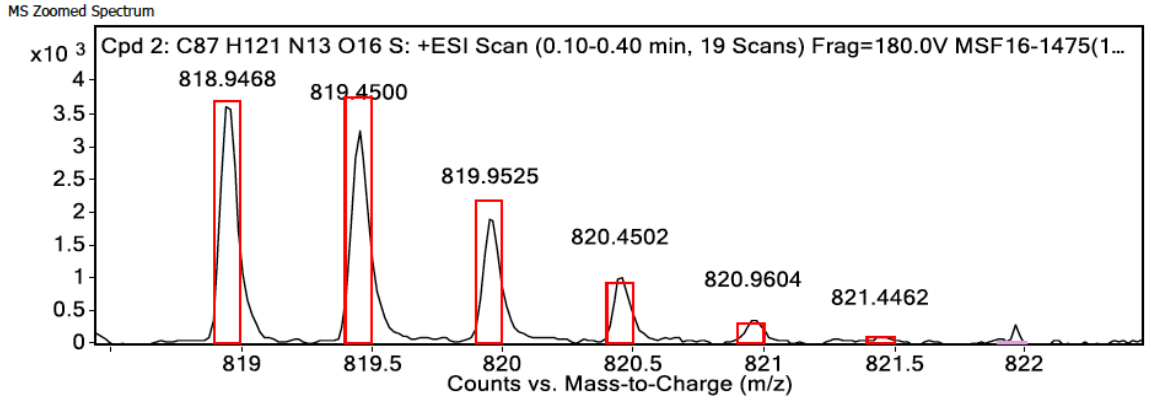
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
365.10700			565769.84			
847.97030	847.96700	2	25008.84	C90H127N13O17S	(M+2H)+2	-3.97
848.47240	848.46850	2	24557.61	C90H127N13O17S	(M+2H)+2	-4.59
848.97370	848.96960	2	14652.48	C90H127N13O17S	(M+2H)+2	-4.85
849.47350	849.47060	2	6741.03	C90H127N13O17S	(M+2H)+2	-3.44
849.97120	849.97150	2	2997.97	C90H127N13O17S	(M+2H)+2	0.26
850.45420	850.47240	2	1472.07	C90H127N13O17S	(M+2H)+2	21.37
850.94010	850.97340	2	1058.37	C90H127N13O17S	(M+2H)+2	39.05
851.38800	851.47440	2	921.43	C90H127N13O17S	(M+2H)+2	101.44

Figure 3.64: HRMS data for peptide 3.28. Starting amount: 7.8 x 10^{-4} mmole. Yield: 69%.



m/z: 818.95 (100.0%), 819.45 (94.1%), 819.95 (42.8%), 820.45 (11.7%), 819.44 (4.8%), 819.94 (4.5%), 819.95 (4.5%), 820.45 (4.3%), 819.95 (3.3%), 820.45 (3.1%), 820.95
ac-ala-leu-tyr-leu-val-cys(Atto647N)-gly-glu



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
818.94680	818.94600	2	3749.23	C87H121N13O16S	(M+2H)+2	-0.95
819.45000	819.44760	2	3244.27	C87H121N13O16S	(M+2H)+2	-3
819.95250	819.94870	2	1963.04	C87H121N13O16S	(M+2H)+2	-4.64
820.45020	820.44960	2	1035.57	C87H121N13O16S	(M+2H)+2	-0.78
820.96040	820.95050	2	387.35	C87H121N13O16S	(M+2H)+2	-12.11
821.44620	821.45140	2	111	C87H121N13O16S	(M+2H)+2	6.34
830.43990			3698.11			

Figure 3.65: HRMS data for peptide 3.29. Starting amount: 8.2 x 10^{-4} mmole. Yield: 88%

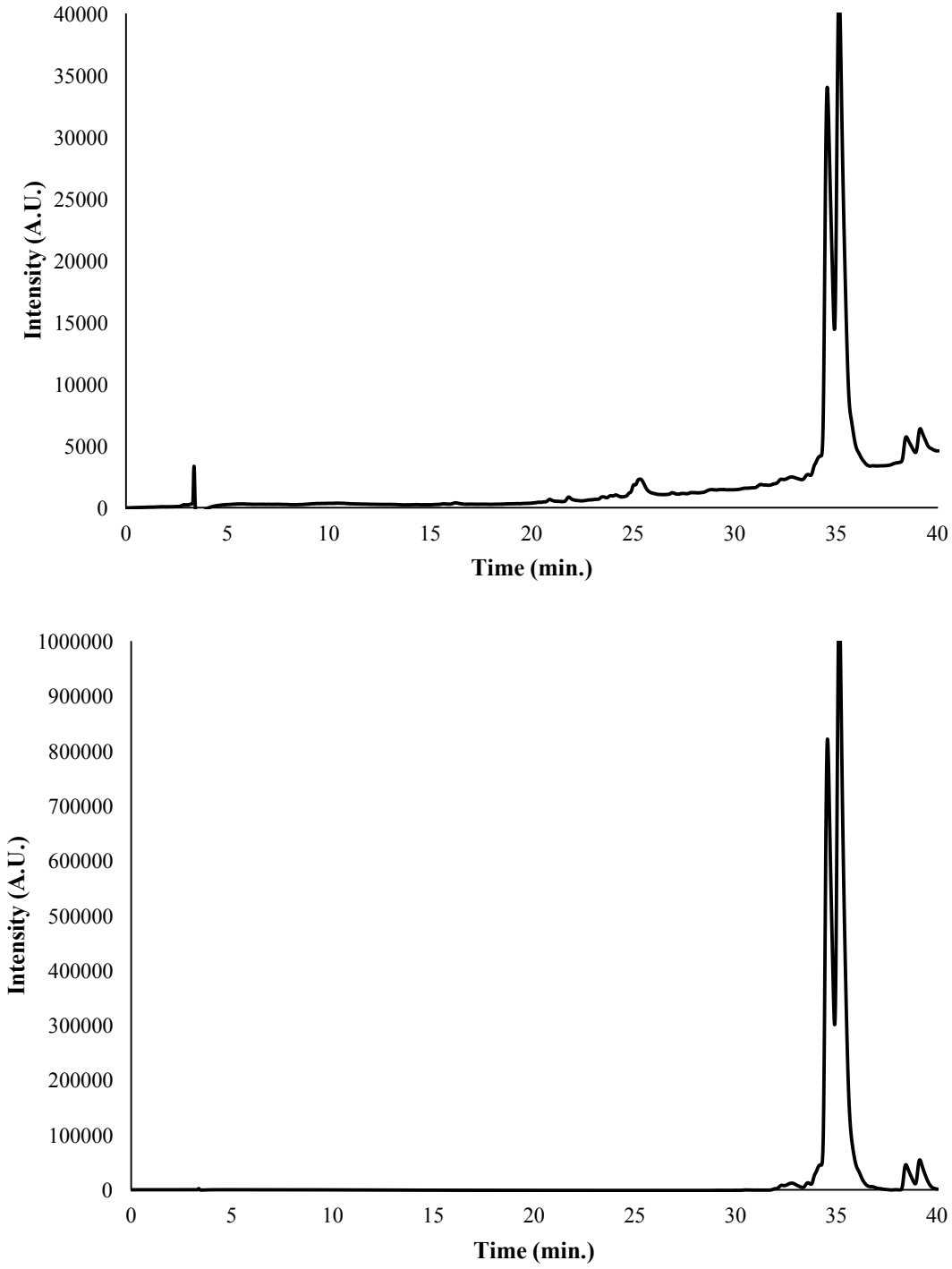
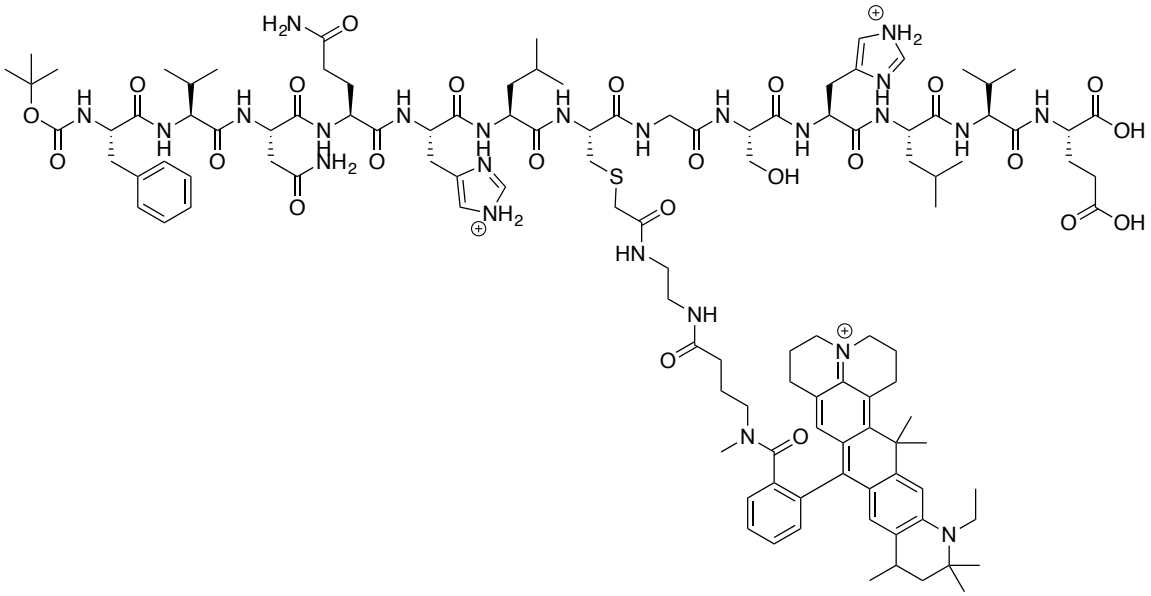
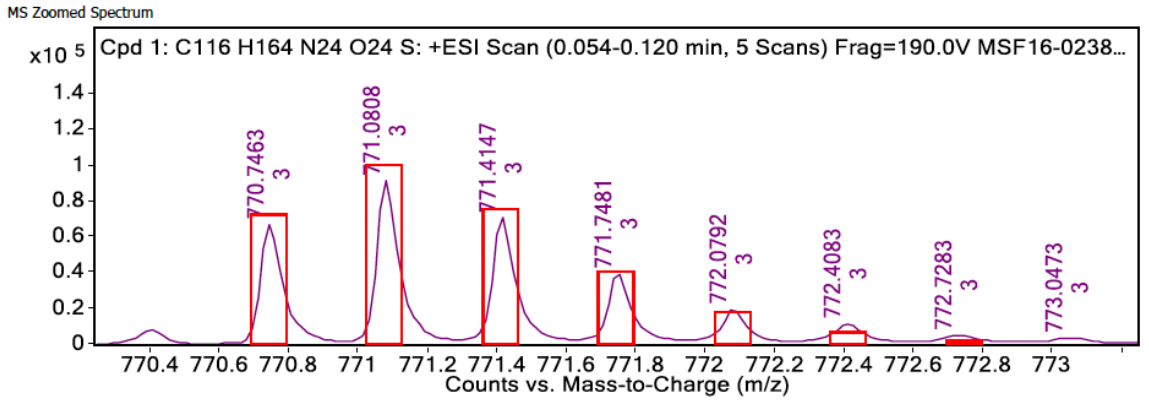


Figure 3.66: (top) Analytical trace for peptide 3.29 at 254 nm; Retention times: 34.51733 min., 35.08267 min. (bottom) Analytical trace at 640 nm; Retention times: 34.51733 min., 35.08267 min.



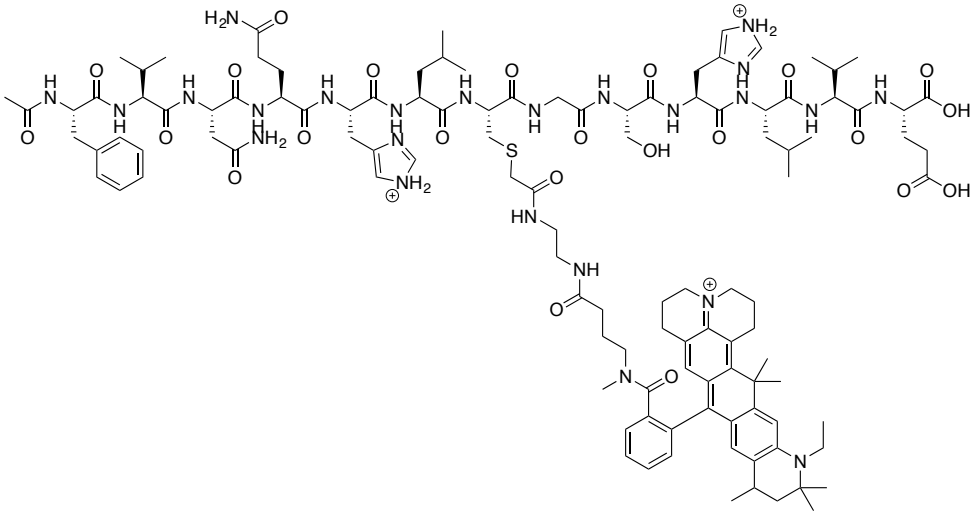
m/z: 771.08 (100.0%), 770.74 (79.7%), 771.41 (34.1%), 771.41 (28.1%), 771.75 (21.4%), 771.41 (8.9%), 772.08 (7.4%), 771.08 (7.1%), 771.74 (5.5%), 771.75 (4.9%), 771.74 (4.5%), 771.75 (4.2%), 771.41 (3.9%), 771.41 (3.6%), 772.08 (3.1%), 771.41 (1.9%), 772.42 (1.9%), 772.08 (1.5%), 771.08 (1.5%), 772.08 (1.4%), 772.08 (1.3%), 772.41 (1.1%)



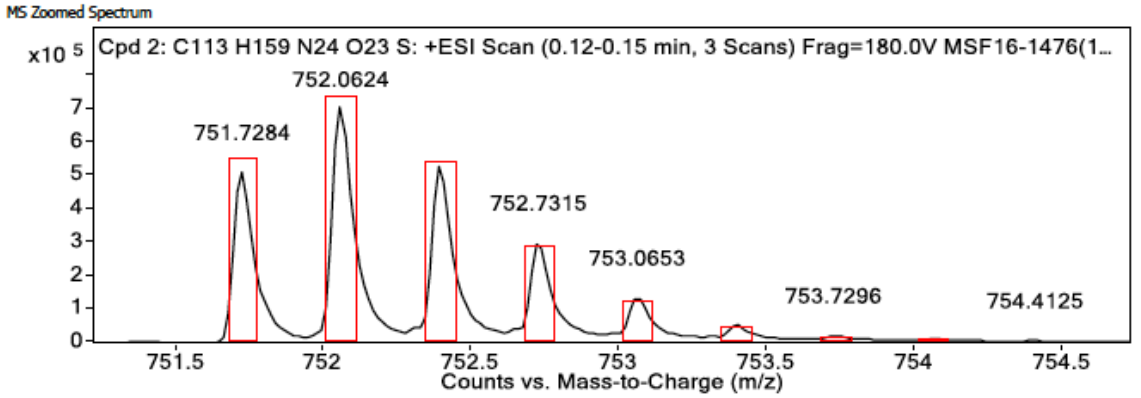
MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
365.10550			140897.64			
770.74630	770.74300	3	67632.23	C116H164N24O24S	(M+3H)+3	-4.3
771.08080	771.07730	3	91984.34	C116H164N24O24S	(M+3H)+3	-4.54
771.41470	771.41150	3	71127.85	C116H164N24O24S	(M+3H)+3	-4.26
771.74810	771.74550	3	39947.97	C116H164N24O24S	(M+3H)+3	-3.3
772.07920	772.07950	3	20175.87	C116H164N24O24S	(M+3H)+3	0.36
772.40830	772.41350	3	12062.53	C116H164N24O24S	(M+3H)+3	6.76
772.72830	772.74750	3	5683.24	C116H164N24O24S	(M+3H)+3	24.82
773.04730	773.08150	3	4204.53	C116H164N24O24S	(M+3H)+3	44.23

Figure 3.67: HRMS data for peptide 3.30. Starting amount: 9.4 x 10^{-4} mmole. Yield: 54%.



m/z: 751.73 (100.0%), 751.40 (81.8%), 752.06 (35.0%), 752.06 (25.6%), 752.40 (20.8%), 752.06 (8.9%), 751.73 (7.3%), 752.73 (6.9%), 752.40 (5.4%), 752.40 (4.5%), ac-phe-val-asn-gln-his-leu-cys(Atto647N)-gly-ser-his-leu-val-glu



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
751.72840	751.73160	3	511401.06	C113H159N24O23S	(M+3H)+3	4.25
752.06240	752.06600	3	705494.91	C113H159N24O23S	(M+3H)+3	4.76
752.39690	752.40010	3	532292.21	C113H159N24O23S	(M+3H)+3	4.26
752.73150	752.73410	3	300126.69	C113H159N24O23S	(M+3H)+3	3.55
753.06530	753.06810	3	135330.75	C113H159N24O23S	(M+3H)+3	3.83
753.39810	753.40210	3	53984.58	C113H159N24O23S	(M+3H)+3	5.32
753.72960	753.73610	3	21959.77	C113H159N24O23S	(M+3H)+3	8.64
754.07120	754.07010	3	11559.43	C113H159N24O23S	(M+3H)+3	-1.47
754.41250	754.40410	3	6943.85	C113H159N24O23S	(M+3H)+3	-11.09
754.70930	754.73820	3	4510.95	C113H159N24O23S	(M+3H)+3	38.24

Figure 3.68: HRMS data for peptide 3.31. Starting amount: 9.7 x 10^{-4} mmole. Yield: 86%.

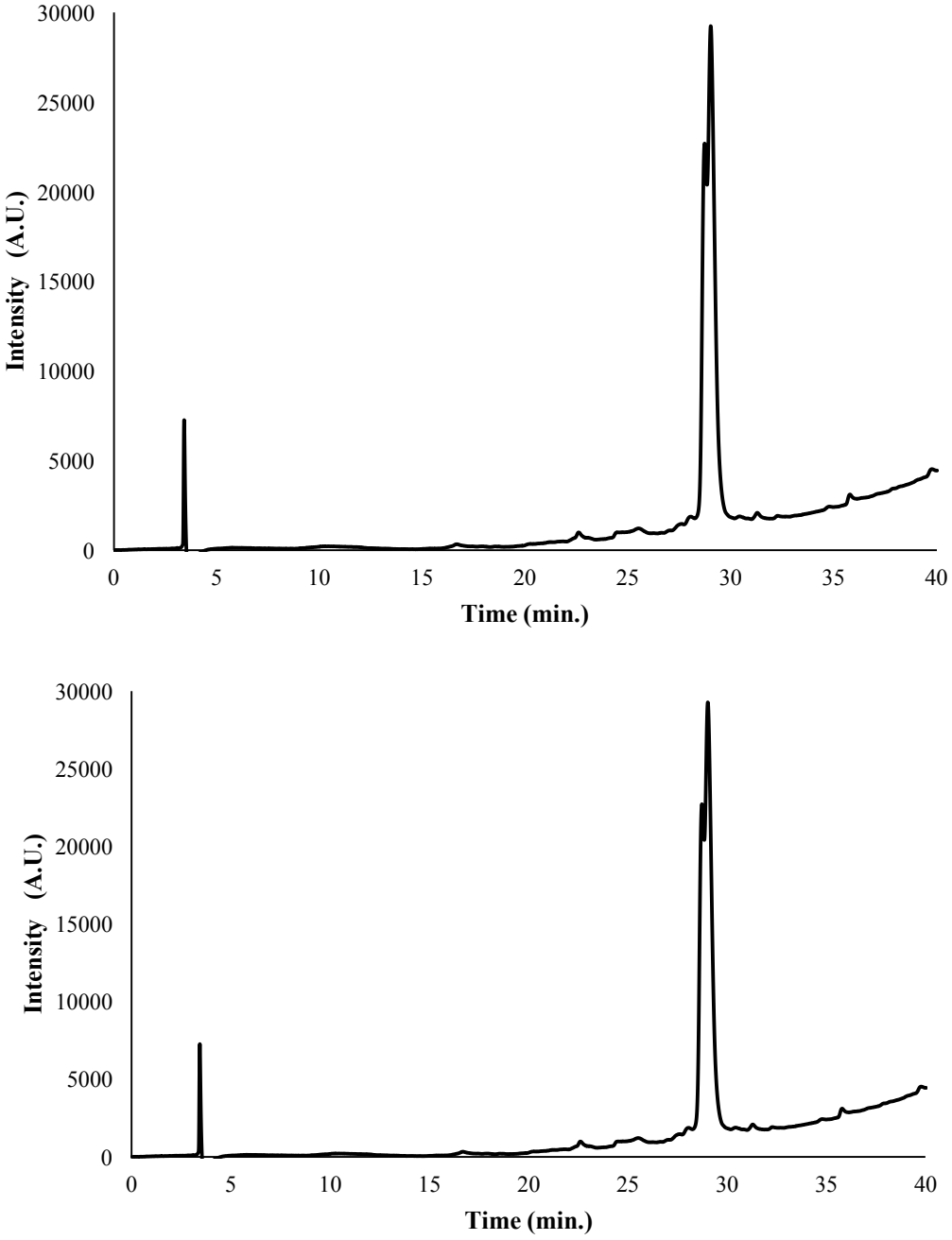
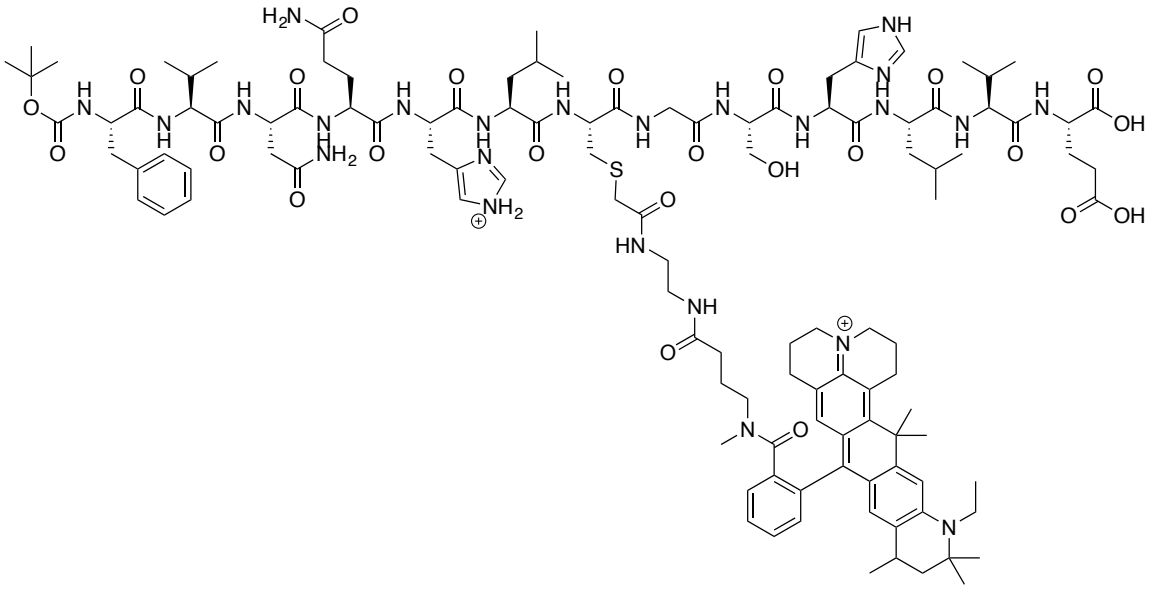
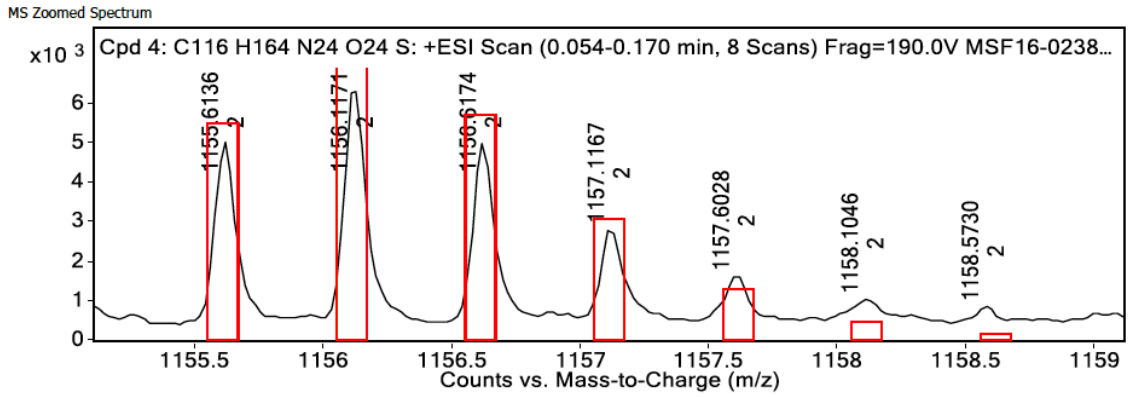


Figure 3.69: (top) Analytical trace for peptide 3.31 at 254 nm. Retention times: 28.68267 min., 28.992 min. (bottom) Analytical trace at 640 nm; Retention times: 28.6862 min., 25.992 min.



m/z: 1156.11 (100.0%), 1155.61 (79.7%), 1156.61 (62.2%), 1157.12 (16.2%), 1157.12 (9.3%), 1156.61 (8.9%), 1156.11 (7.1%), 1157.62 (5.3%), 1157.11 (4.9%), 1157.11 (4.5%), 1156.61 (3.9%), 1156.61 (3.6%), 1157.62 (3.1%), 1157.11 (3.0%), 1157.61 (2.8%), 1157.11 (2.5%), 1157.62 (2.5%), 1156.62 (1.9%), 1157.61 (1.9%), 1158.12 (1.6%), 1156.11 (1.5%), 1157.12 (1.2%), 1158.11 (1.2%)

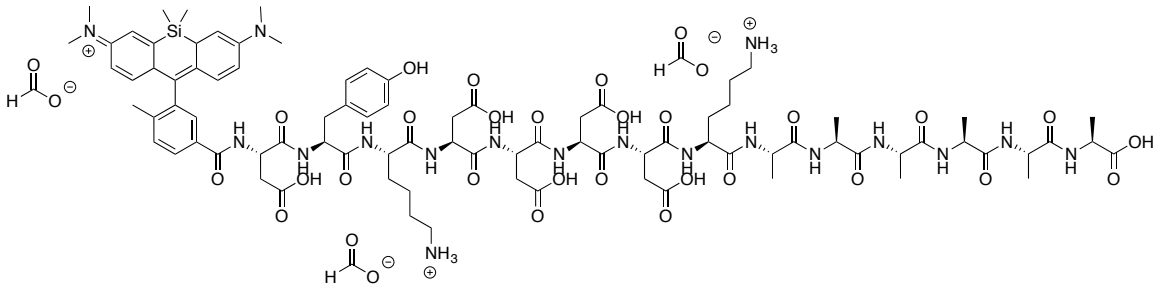
boc-phe-val-asn-gln-his-leu-cys(Atto647N)-gly-ser-his-leu-val-glu

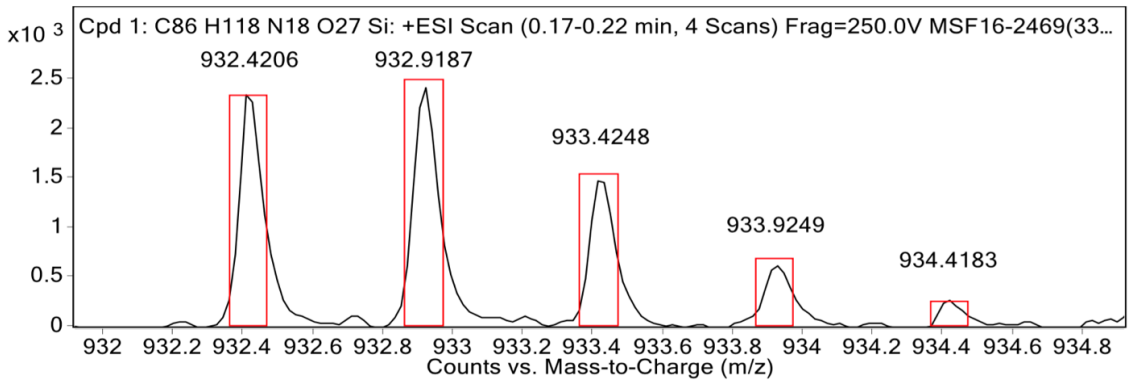


MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
365.10580			100408.45			
1155.61360	1155.61080	2	5053.2	C116H164N24O24S	(M+2H)+2	-2.42
1156.11710	1156.11230	2	6474.78	C116H164N24O24S	(M+2H)+2	-4.15
1156.61740	1156.61350	2	5008.68	C116H164N24O24S	(M+2H)+2	-3.33
1157.11670	1157.11460	2	2864.8	C116H164N24O24S	(M+2H)+2	-1.81
1157.60280	1157.61560	2	1656.22	C116H164N24O24S	(M+2H)+2	11.13
1158.10460	1158.11660	2	1080.67	C116H164N24O24S	(M+2H)+2	10.42
1158.57300	1158.61760	2	884.34	C116H164N24O24S	(M+2H)+2	38.53
1159.03790	1159.11870	2	727.88	C116H164N24O24S	(M+2H)+2	69.65

Figure 3.70: HRMS data for peptide 3.32.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
238.10000			20657.5			
932.42060	932.41640	2	2411.28	C86H118N18O27Si	(M+2H)+2	-4.48
932.91870	932.91780	2	2430.01	C86H118N18O27Si	(M+2H)+2	-0.96
933.42480	933.41900	2	1519.54	C86H118N18O27Si	(M+2H)+2	-6.2
933.92490	933.92000	2	628.83	C86H118N18O27Si	(M+2H)+2	-5.25
934.41830	934.42110	2	275.89	C86H118N18O27Si	(M+2H)+2	3

Figure 3.71: HRMS data for peptide 3.33. Starting amount: 0.012 mmole. Yield: 1.4%.

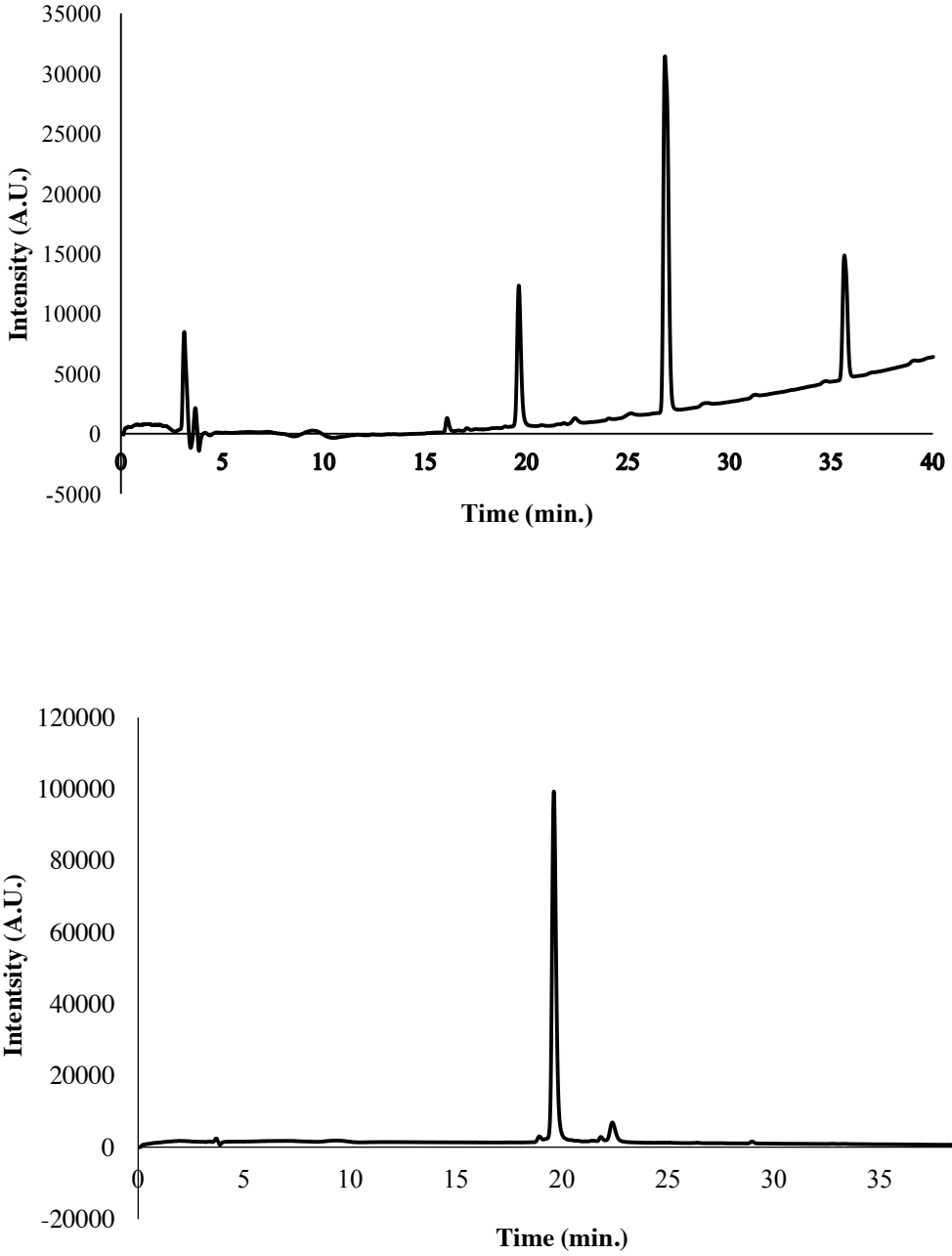
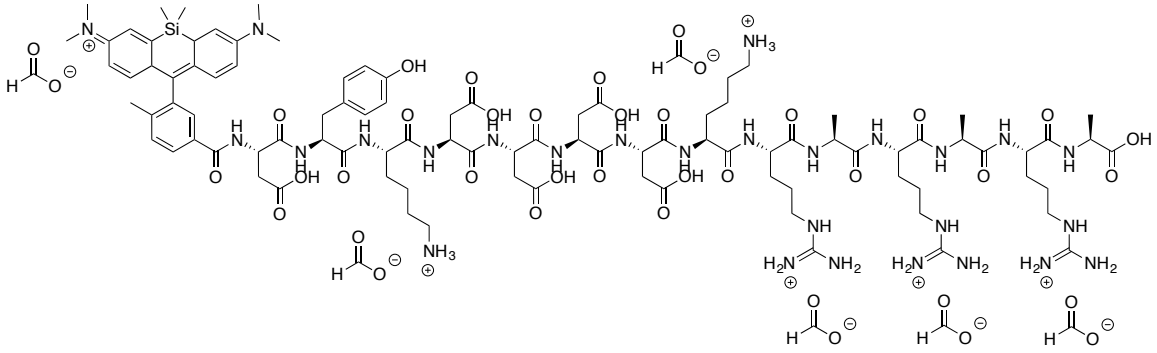
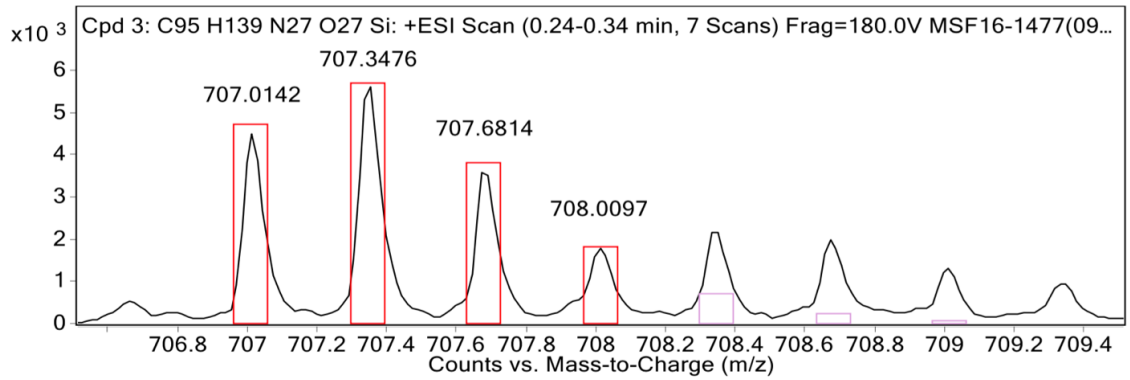


Figure 3.72: (top) Analytical trace for peptide 3.33 at 254 nm; Retention times: 19.57333 min., 26.77333 min., 35.59467 min. (bottom) Analytical trace at 640 nm; Retention times: 19.584 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
707.01420	707.01070	3	4516.93	C95H139N27O27Si	(M+3H)+3	-4.95
707.34760	707.34500	3	5694.75	C95H139N27O27Si	(M+3H)+3	-3.69
707.68140	707.67900	3	3718.89	C95H139N27O27Si	(M+3H)+3	-3.35
708.00970	708.01310	3	1813	C95H139N27O27Si	(M+3H)+3	4.79
959.96500			76608.47			

Figure 3.73: HRMS data for peptide 3.34. Starting amount: 0.012 mmole. Yield: 4.1%.

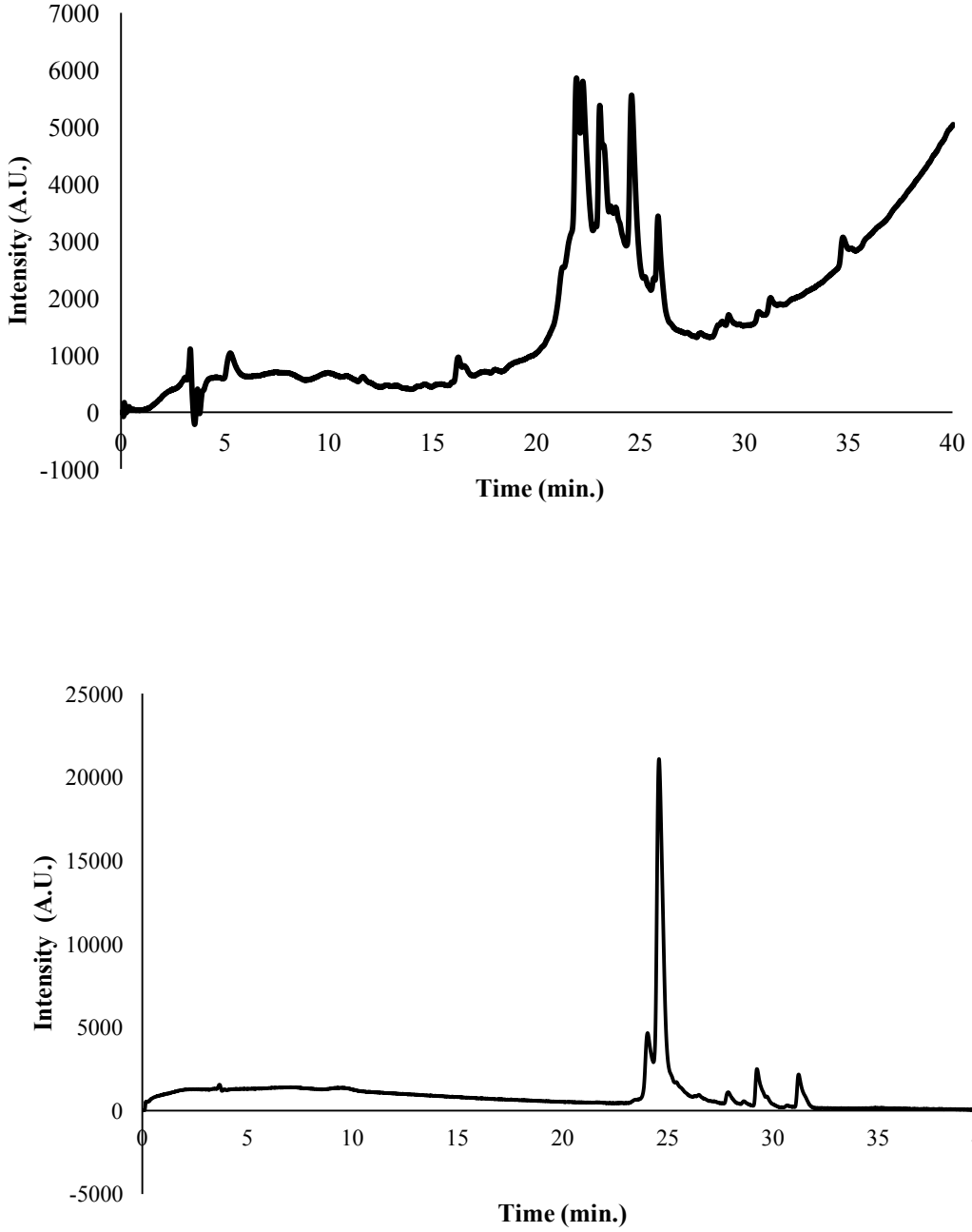
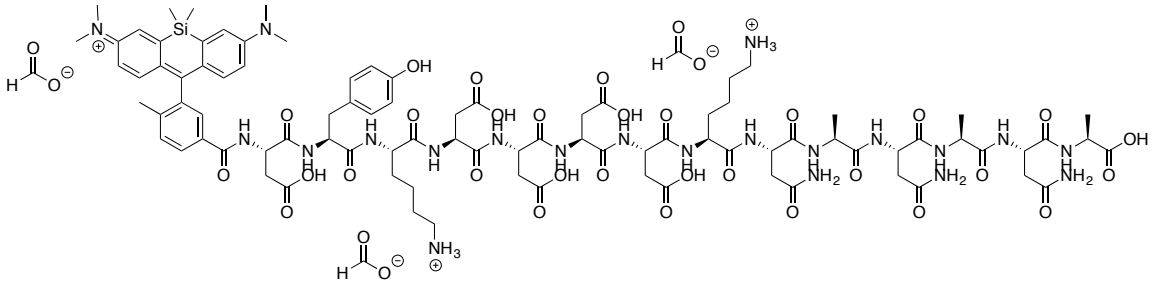
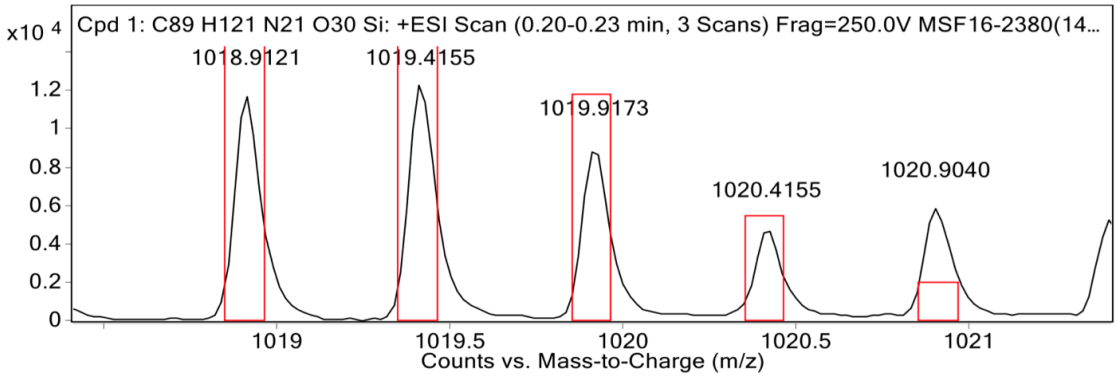


Figure 3. 74: (top) Analytical trace for peptide 3.34 at 254 nm; Retention times: 21.866 min., 21.952 min., 22.99733 min., 24.512 min., 25.792 min. (bottom) Analytical trace at 640 nm; Retention time: 24.544 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
687.27520			47138.57			
1018.91210	1018.90710	2	11794.48	C89H121N21O30Si	(M+2Na)+2	-4.93
1019.41550	1019.40850	2	12420.73	C89H121N21O30Si	(M+2Na)+2	-6.92
1019.91730	1019.90960	2	9098.96	C89H121N21O30Si	(M+2Na)+2	-7.48
1020.41550	1020.41070	2	4826.24	C89H121N21O30Si	(M+2Na)+2	-4.76
1020.90400	1020.91170	2	5870.43	C89H121N21O30Si	(M+2Na)+2	7.53
1021.40390	1021.41280	2	5338.25	C89H121N21O30Si	(M+2Na)+2	8.64
1021.90550	1021.91380	2	4005.9	C89H121N21O30Si	(M+2Na)+2	8.17
1022.40190	1022.41490	2	1935.6	C89H121N21O30Si	(M+2Na)+2	12.7

Figure 3.75: HRMS data for peptide 3.35. Starting amount: 0.012 mmole. Yield: 2.5%.

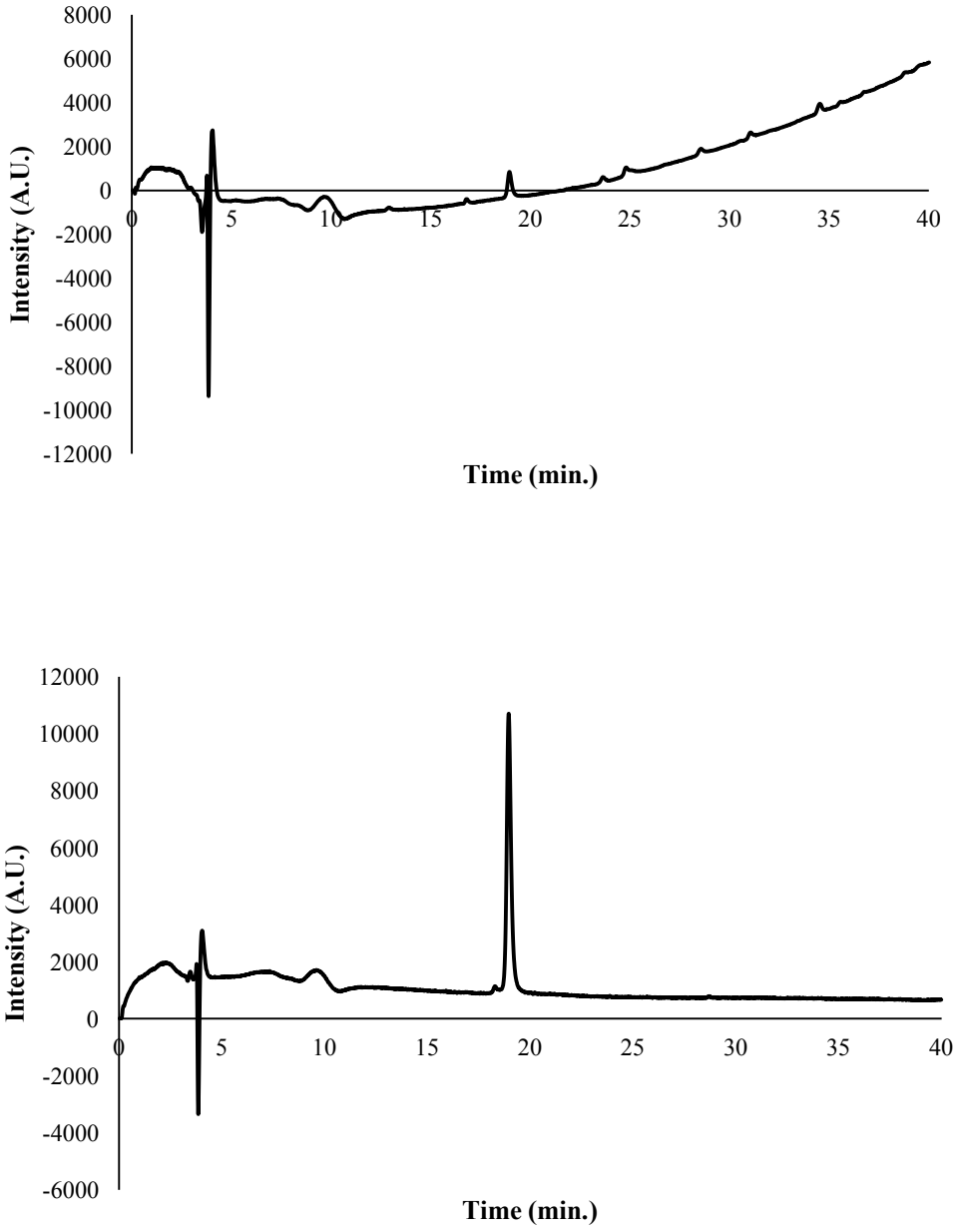
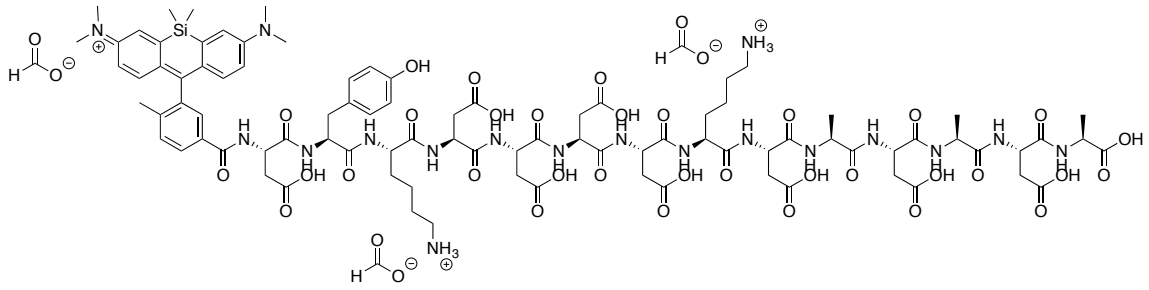
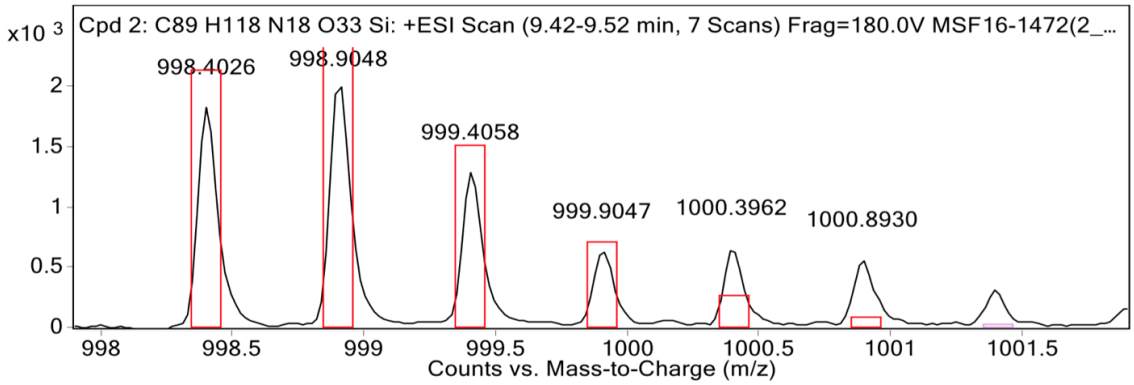


Figure 3.76: (top) Analytical trace for peptide 3.35 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 18.92267 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
666.27540			75024.1			
998.40260	998.40120	2	1838.32	C89H118N18O33Si	(M+2H)+2	-1.44
998.90480	998.90260	2	2038.57	C89H118N18O33Si	(M+2H)+2	-2.22
999.40580	999.40370	2	1299.09	C89H118N18O33Si	(M+2H)+2	-2.08
999.90470	999.90480	2	641.74	C89H118N18O33Si	(M+2H)+2	0.14
1000.39620	1000.40590	2	671.35	C89H118N18O33Si	(M+2H)+2	9.63
1000.89300	1000.90690	2	573.16	C89H118N18O33Si	(M+2H)+2	13.87

Figure 3.77: HRMS data for peptide 3.36. Starting amount: 0.012 mmole. Yield: 3.4%.

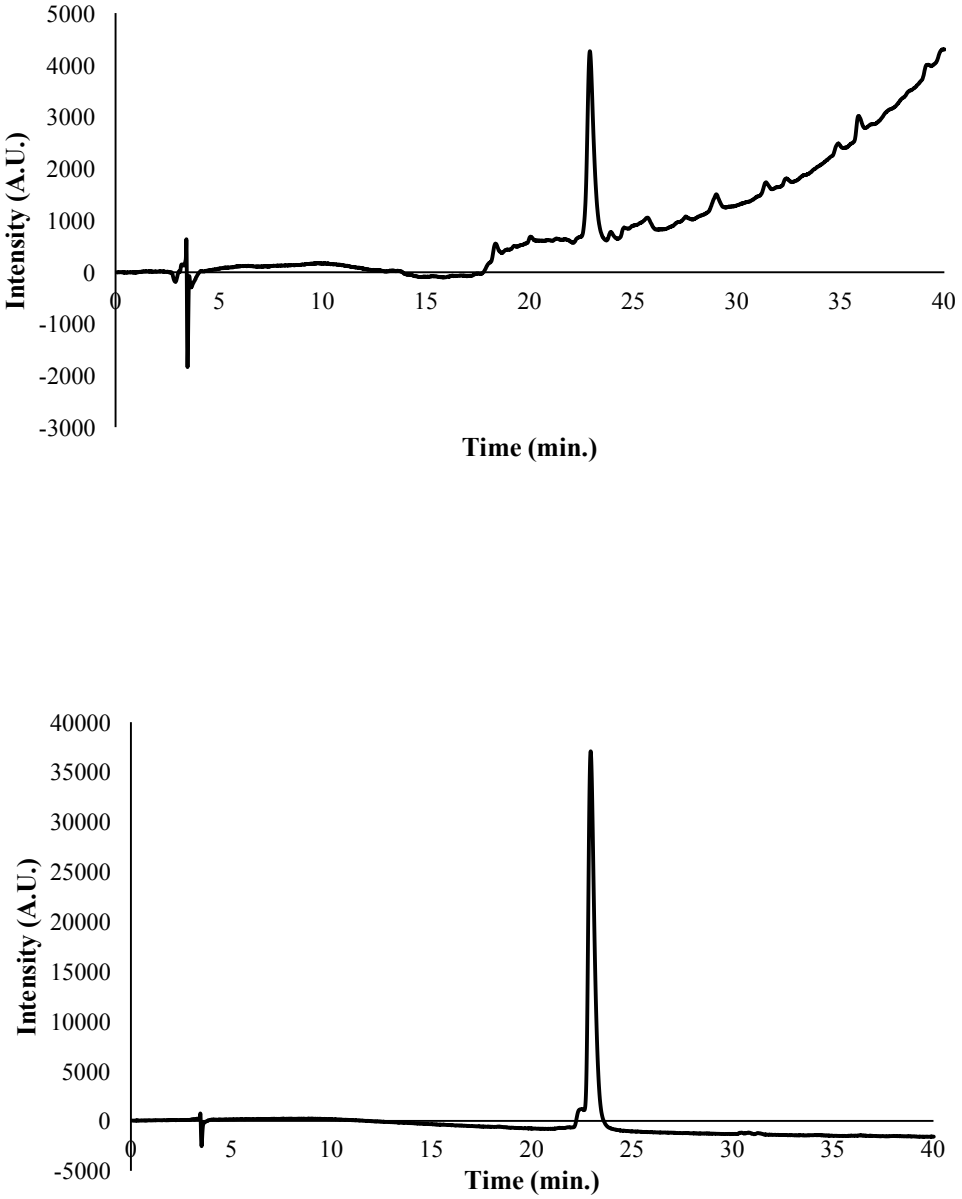
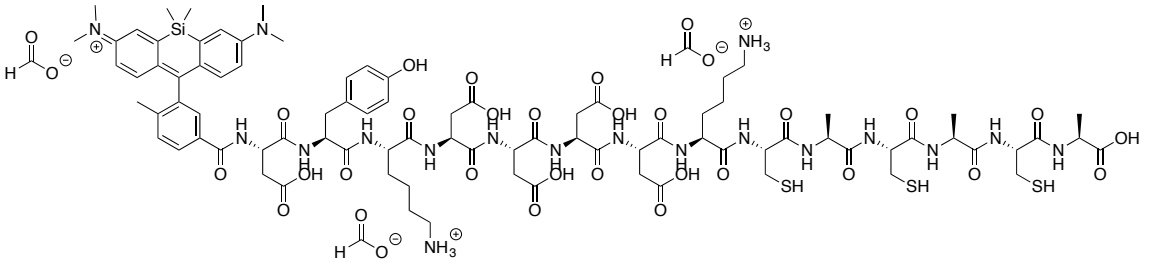
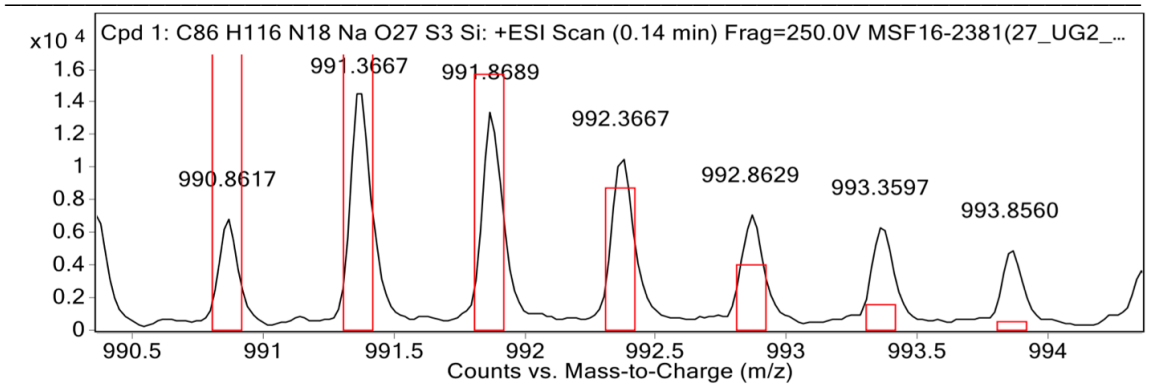


Figure 3.78: (top) Analytical trace for peptide 3.36 at 254 nm; Retention time: 22.85867 min. (bottom) Analytical trace at 640 nm; Retention time: 22.88 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
990.86170	990.86160	2	6920.15	C86H116N18NaO27S3Si	(M+2H)+2	-0.07
991.36670	991.36300	2	15069.25	C86H116N18NaO27S3Si	(M+2H)+2	-3.72
991.86890	991.86330	2	13490.69	C86H116N18NaO27S3Si	(M+2H)+2	-5.66
992.36670	992.36380	2	10630.63	C86H116N18NaO27S3Si	(M+2H)+2	-2.97
992.86290	992.86410	2	7181.26	C86H116N18NaO27S3Si	(M+2H)+2	1.21
993.35970	993.36440	2	6405.23	C86H116N18NaO27S3Si	(M+2H)+2	4.76

Figure 3.79: HRMS data for peptide 3.37. Starting amount: 0.012 mmole. Yield: 1.9%.

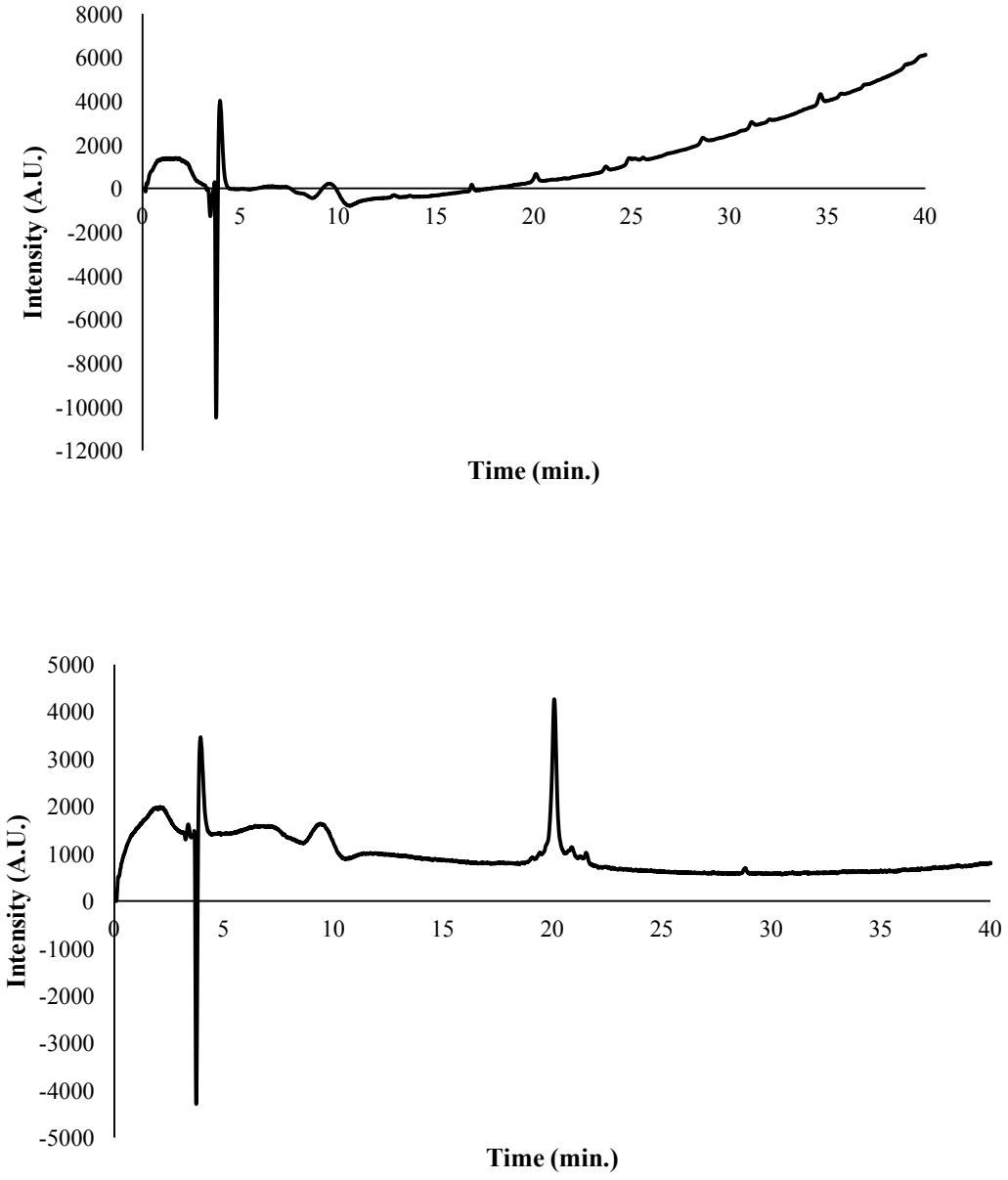
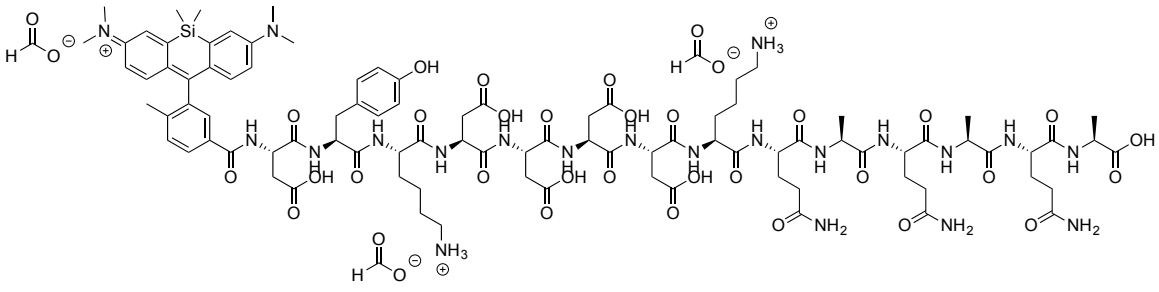
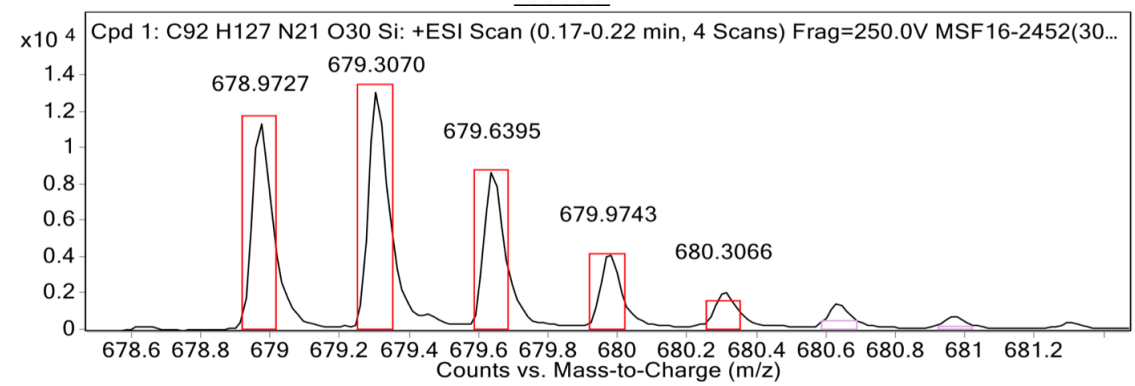


Figure 3.80: (top) Analytical trace for peptide 3.37 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 20.05333 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
678.97270	678.96820	3	11398.4	C92H127N21O30Si	(M+3H)+3	-6.61
679.30700	679.30240	3	13117.34	C92H127N21O30Si	(M+3H)+3	-6.75
679.63950	679.63660	3	8794.98	C92H127N21O30Si	(M+3H)+3	-4.4
679.97430	679.97060	3	4217.21	C92H127N21O30Si	(M+3H)+3	-5.44
680.30660	680.30460	3	2155.24	C92H127N21O30Si	(M+3H)+3	-2.87
959.96570			18432.5			

Figure 3.81: HRMS data for peptide 3.38. Starting amount: 0.012 mmole. Yield: 0.78%.

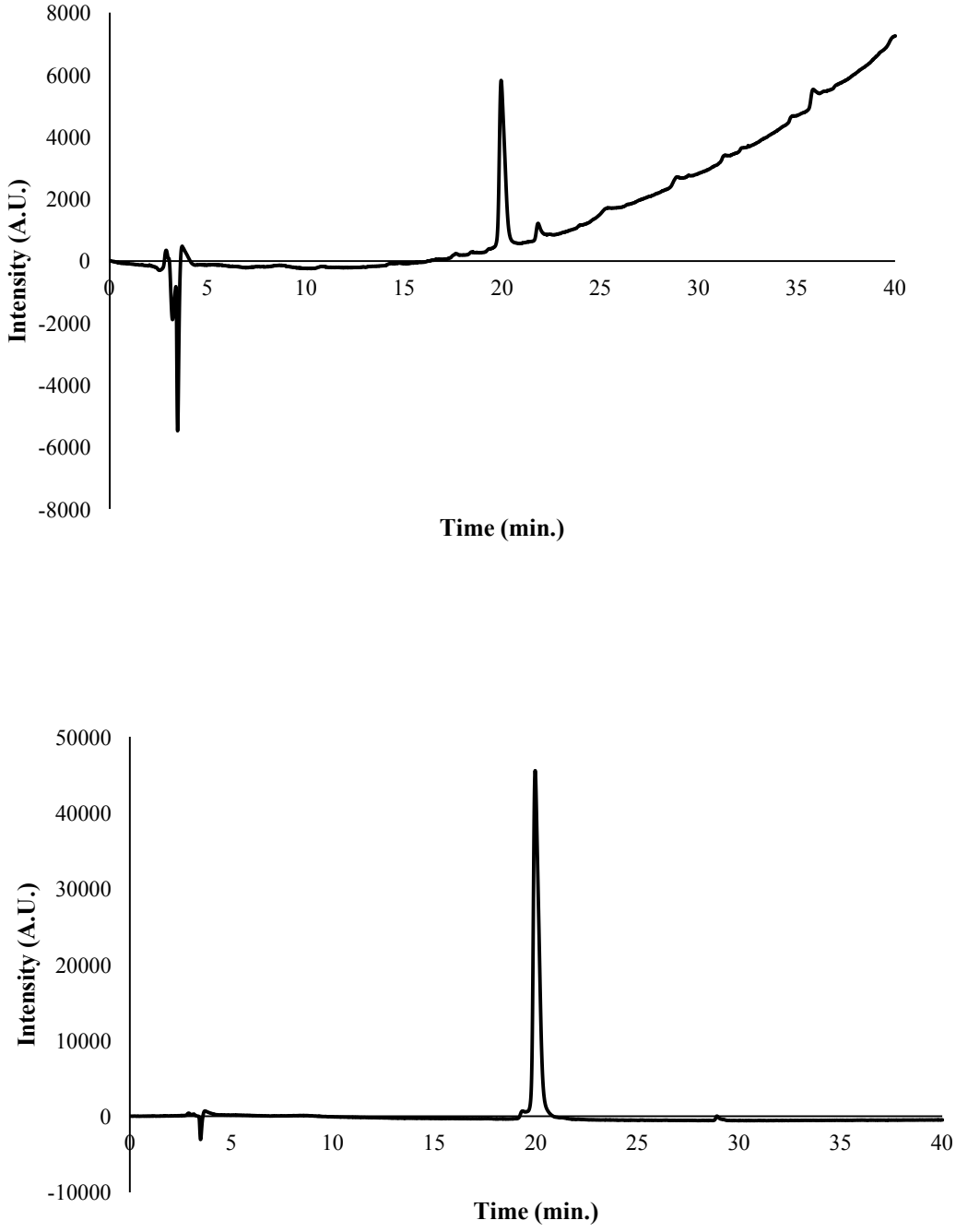
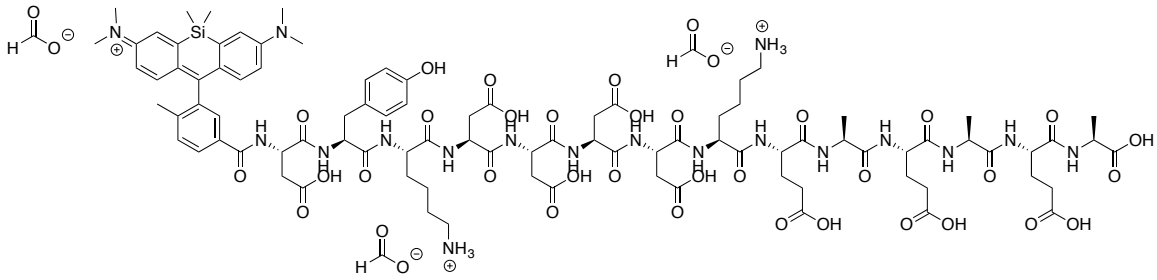
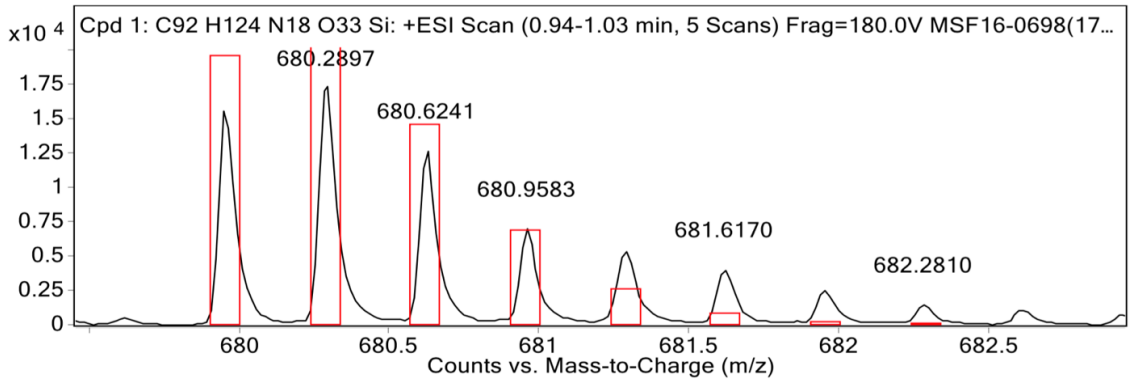


Figure 3.82: (top) Analytical trace for peptide 3.38 at 254 nm; Retention time: 19.91467 min. (bottom) Analytical trace at 640 nm; Retention time: 19.91467 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
679.95500	679.95220	3	15851.21	C92H124N18O33Si	(M+3H)+3	-4.17
680.28970	680.28650	3	17806.39	C92H124N18O33Si	(M+3H)+3	-4.75
680.62410	680.62060	3	12709.27	C92H124N18O33Si	(M+3H)+3	-5.15
680.95830	680.95470	3	7027.04	C92H124N18O33Si	(M+3H)+3	-5.38
681.28610	681.28870	3	5451.6	C92H124N18O33Si	(M+3H)+3	3.74
681.61700	681.62270	3	4076.32	C92H124N18O33Si	(M+3H)+3	8.47
681.95060	681.95680	3	2598.72	C92H124N18O33Si	(M+3H)+3	9.03
682.28100	682.29080	3	1525.68	C92H124N18O33Si	(M+3H)+3	14.44

Figure 3.83: HRMS data for peptide 3.39. Starting amount: 0.012 mmole. Yield: 4.2%.

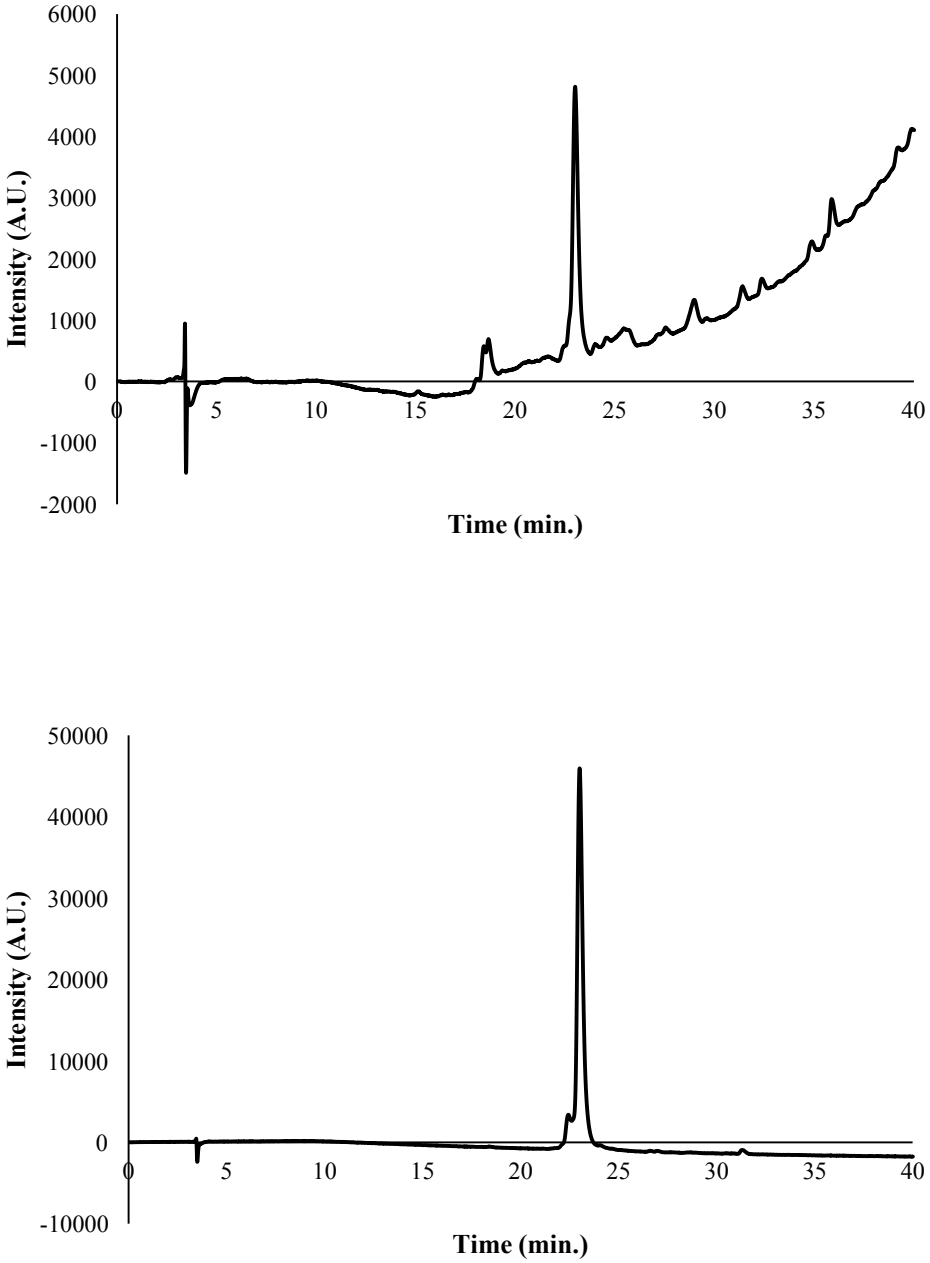
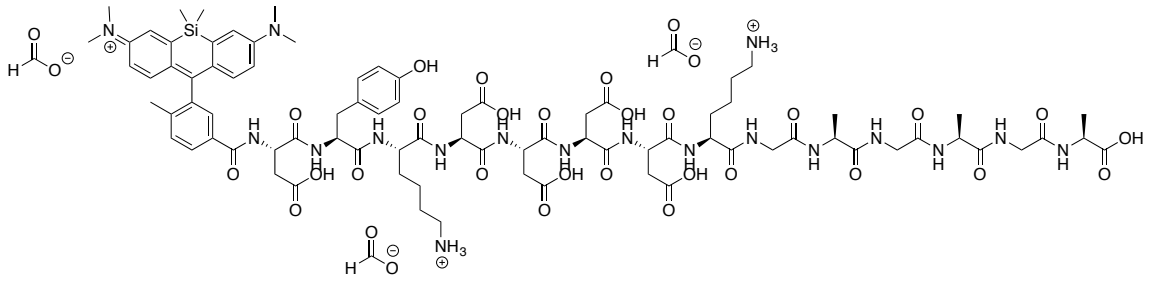
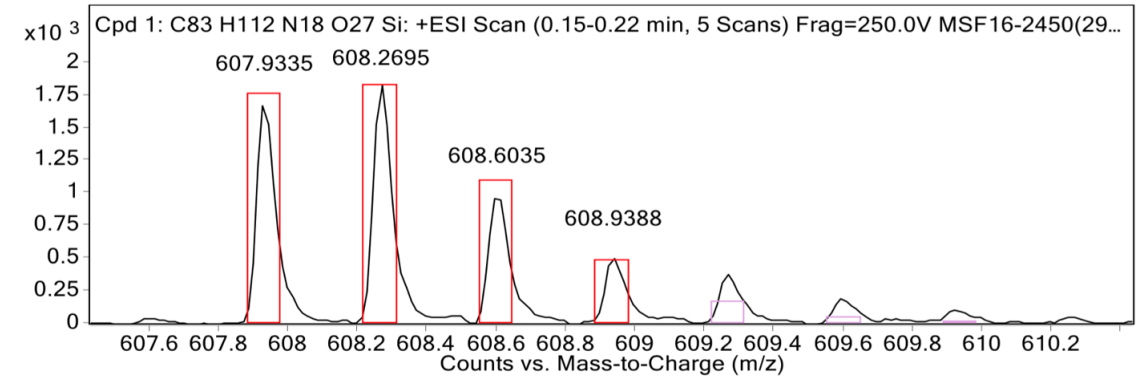


Figure 3.84: (top) Analytical trace for peptide 3.39 at 254 nm; Retention time: 22.96533 min. (bottom) Analytical trace at 640 nm; Retention time: 22.98667 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
607.93350	607.93110	3	1691.62	C83H112N18O27Si	(M+3H)+3	-4.08
608.26950	608.26530	3	1824.04	C83H112N18O27Si	(M+3H)+3	-6.77
608.60350	608.59940	3	986.88	C83H112N18O27Si	(M+3H)+3	-6.8
608.93880	608.93340	3	505.18	C83H112N18O27Si	(M+3H)+3	-8.89
922.01550			142402.2			

Figure 3.85: HRMS data for peptide 3.40. Starting amount: 0.012 mmole. Yield: 0.41%.

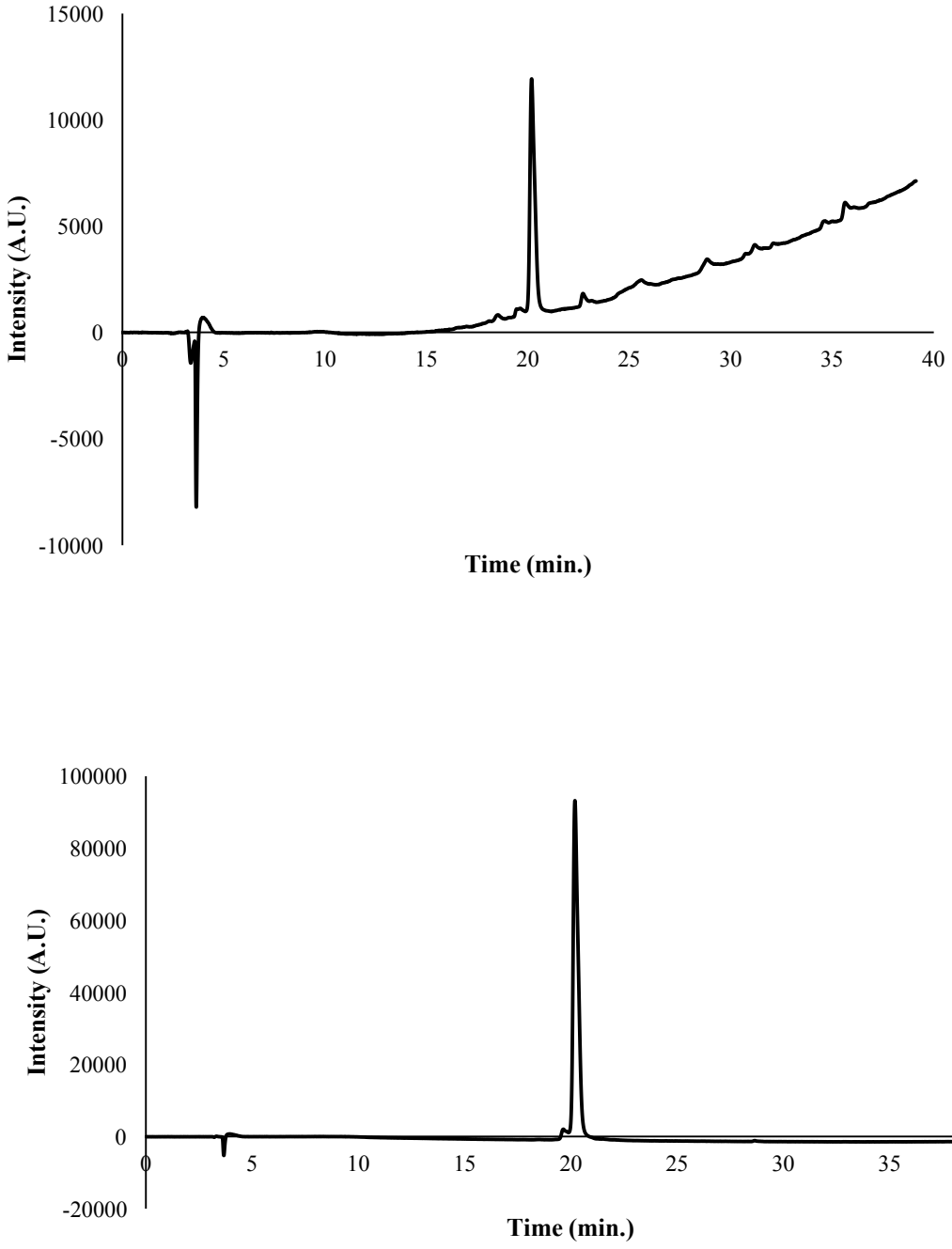
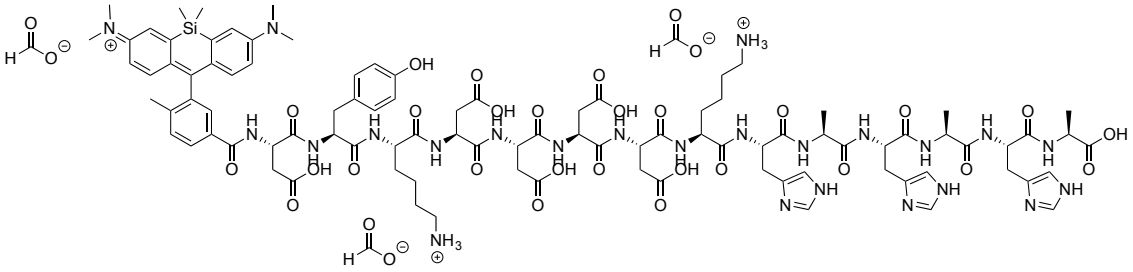
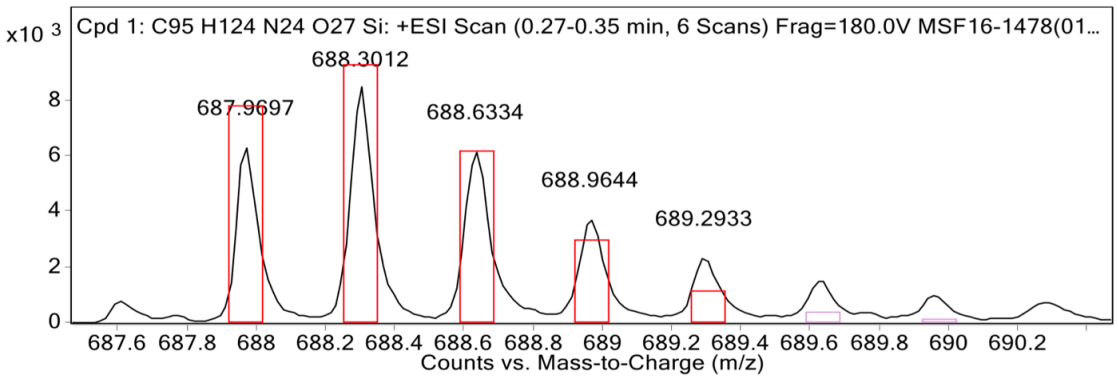


Figure 3.86: (top) Analytical trace for peptide 3.40 at 254 nm; Retention time: 20.16 min. (bottom) Analytical trace at 640 nm; Retention time: 20.16 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
687.96970	687.96850	3	6348.66	C95H124N24O27Si	(M+3H)+3	-1.68
688.30120	688.30280	3	8531.11	C95H124N24O27Si	(M+3H)+3	2.22
688.63340	688.63690	3	6195.78	C95H124N24O27Si	(M+3H)+3	5.02
688.96440	688.97090	3	3752.97	C95H124N24O27Si	(M+3H)+3	9.47
689.29330	689.30490	3	2358.47	C95H124N24O27Si	(M+3H)+3	16.95
959.96520			91754.21			

Figure 3.87: HRMS data for peptide 3.41. Starting amount: 0.012 mmole. Yield: 3.4%.

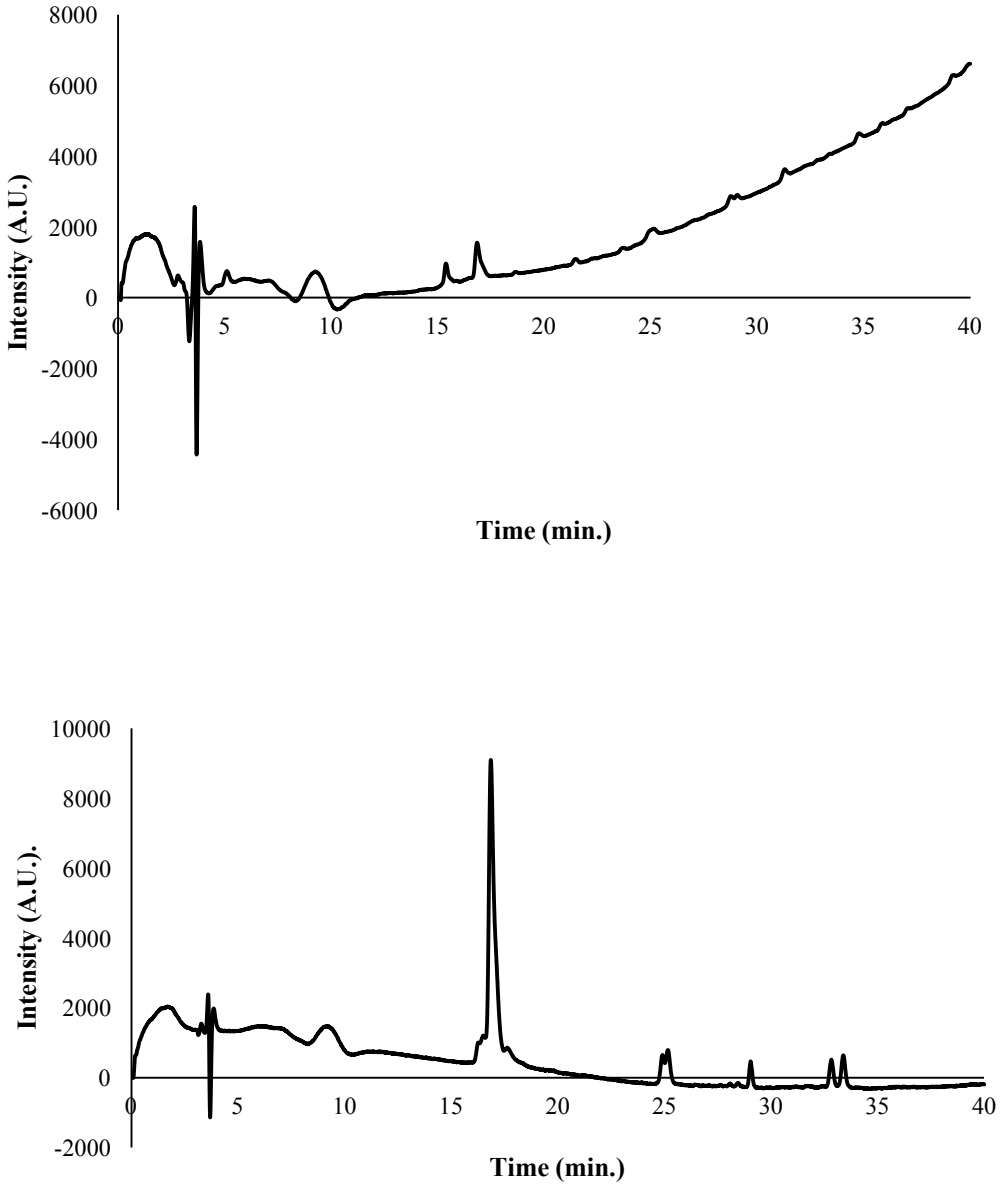
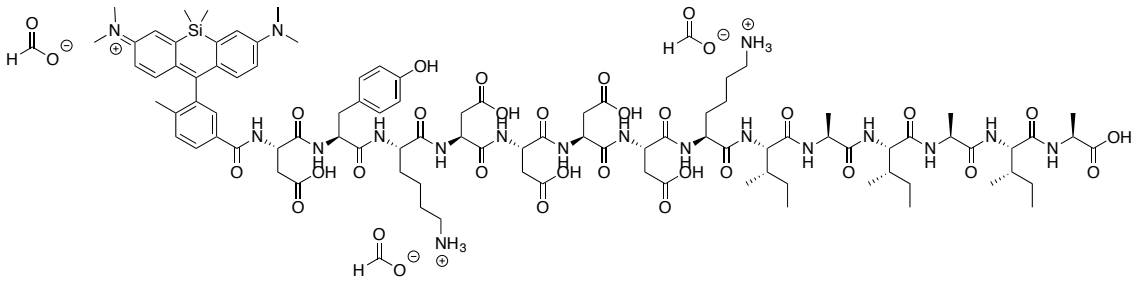
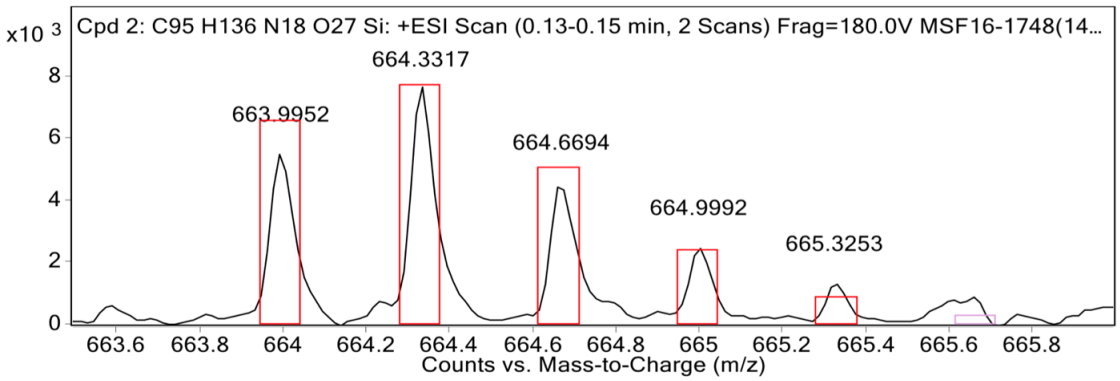


Figure 3.88: (top) Analytical trace for peptide 3.41 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 16.84267 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
663.99520	663.99370	3	5527.52	C95H136N18O27Si	(M+3H)+3	-2.25
664.33170	664.32800	3	7698.32	C95H136N18O27Si	(M+3H)+3	-5.57
664.66940	664.66210	3	4620.34	C95H136N18O27Si	(M+3H)+3	-10.95
664.99920	664.99620	3	2511.81	C95H136N18O27Si	(M+3H)+3	-4.54
665.32530	665.33020	3	1350.36	C95H136N18O27Si	(M+3H)+3	7.33
943.99210			50761.67			

Figure 3.89: HRMS data for peptide 3.42. Starting amount: 0.012 mmole. Yield: 5.2%.

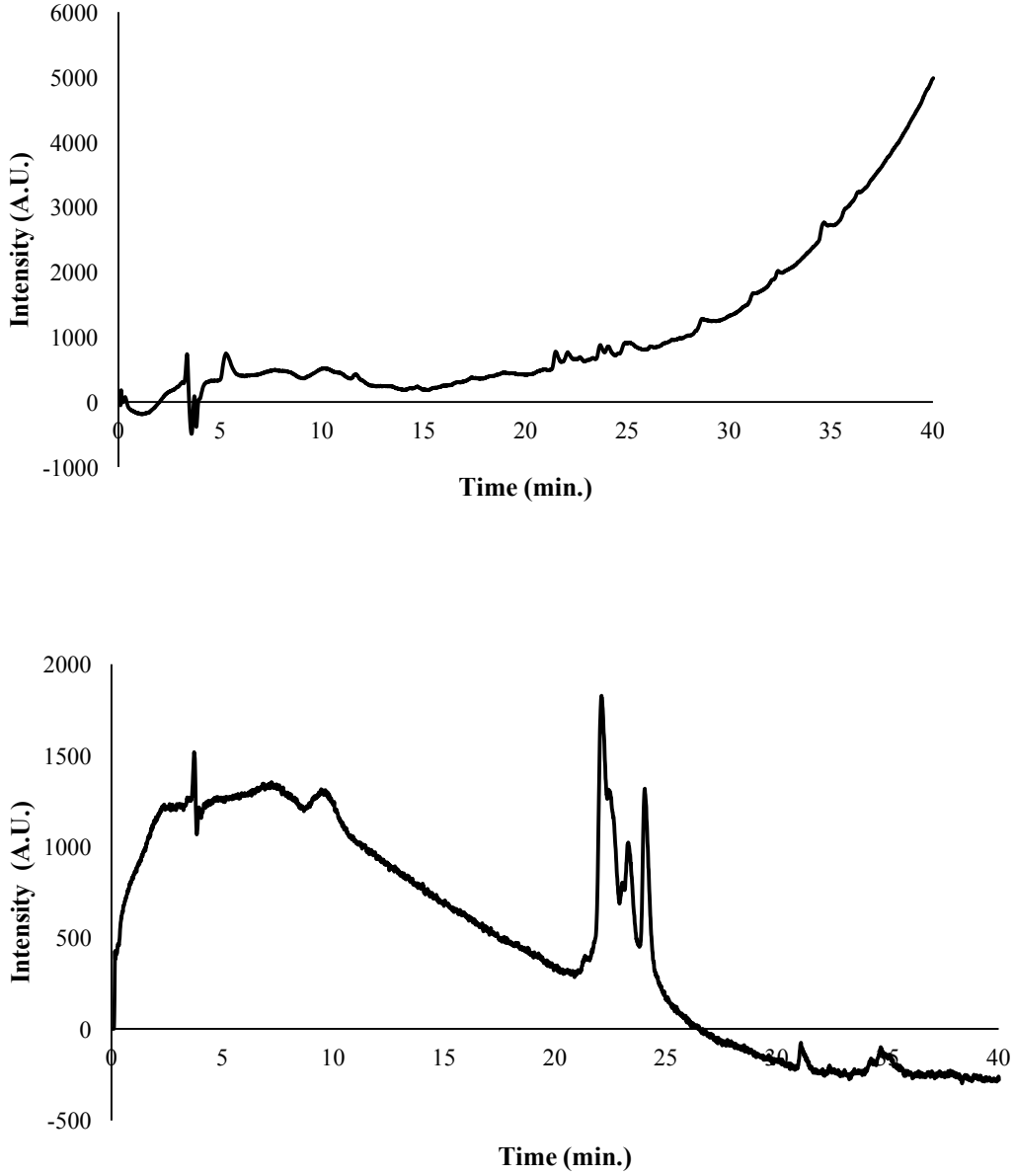
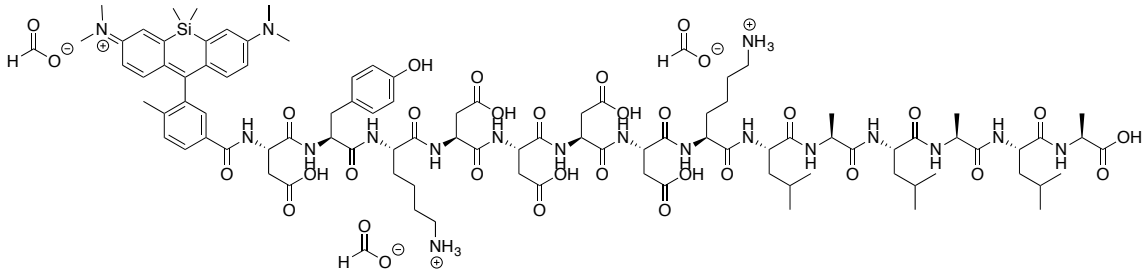
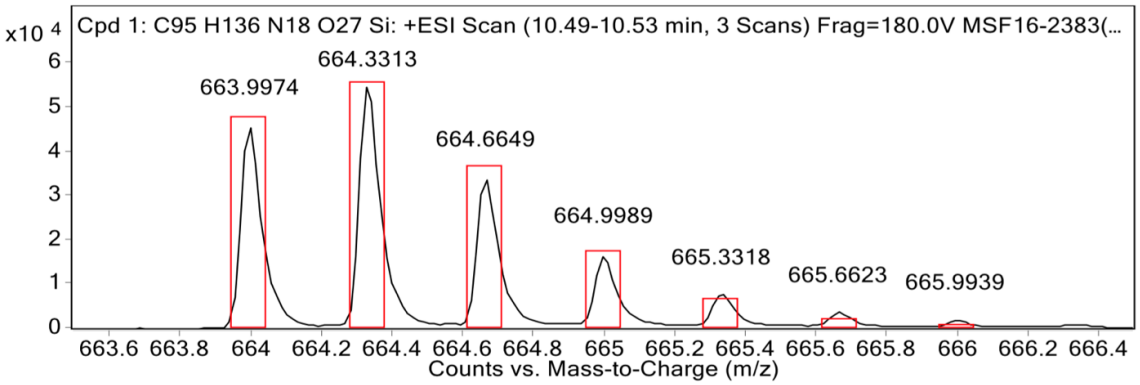


Figure 3.90: (top) Analytical trace for peptide 3.42 at 254 nm. (bottom) Analytical trace at 640 nm; Retention times: 22.06933 min., 23.232 min., 24 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
663.99740	663.99370	3	45569.57	C95H136N18O27Si	(M+3H)+3	-5.6
664.33130	664.32800	3	55649.58	C95H136N18O27Si	(M+3H)+3	-5.03
664.66490	664.66210	3	33854.07	C95H136N18O27Si	(M+3H)+3	-4.21
664.99890	664.99620	3	16538.33	C95H136N18O27Si	(M+3H)+3	-4.07
665.33180	665.33020	3	7119.84	C95H136N18O27Si	(M+3H)+3	-2.39
665.66230	665.66430	3	3086.75	C95H136N18O27Si	(M+3H)+3	2.97
665.99390	665.99830	3	1471.94	C95H136N18O27Si	(M+3H)+3	6.54

Figure 3.91: HRMS data for peptide 3.43. Starting amount: 0.012 mmole. Yield: 3.6%.

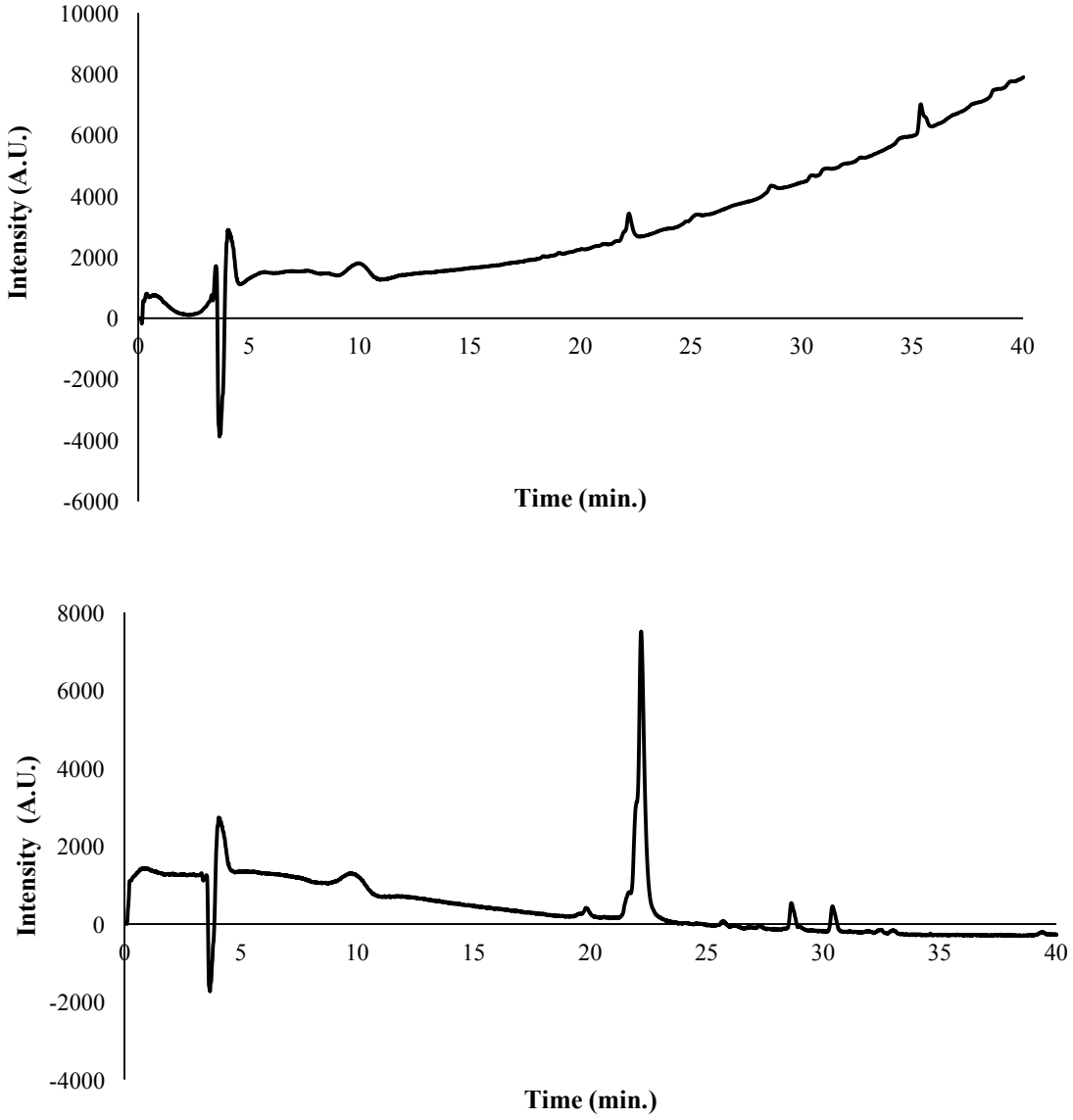
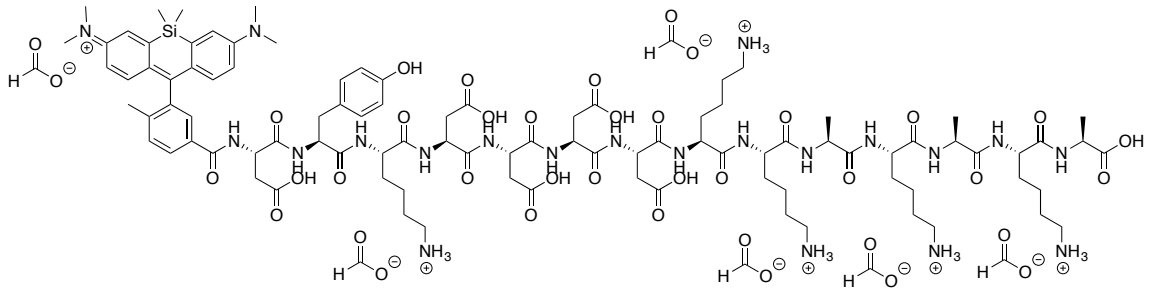
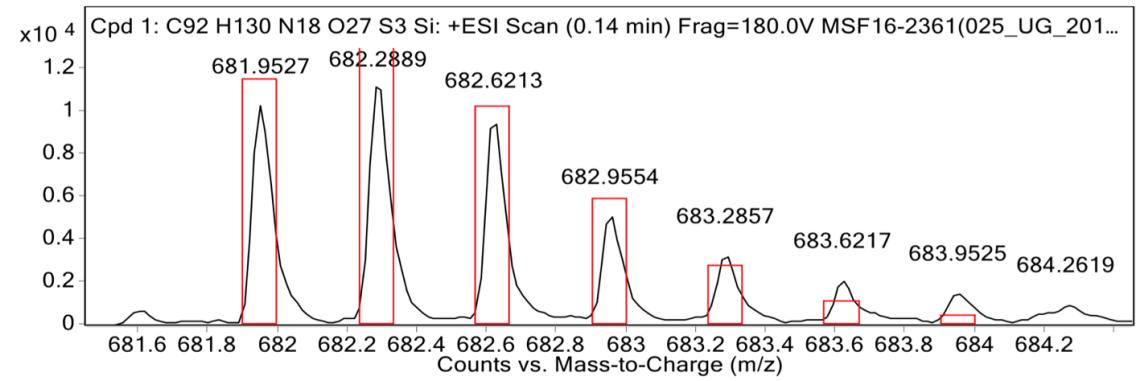


Figure 3.92: (top) Analytical trace for peptide 3.43 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 22.144 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
681.95270	681.95010	3	10246.8	C92H130N18O27S3Si	(M+3H)+3	-3.9
682.28890	682.28430	3	11521	C92H130N18O27S3Si	(M+3H)+3	-6.71
682.62130	682.61800	3	9582.06	C92H130N18O27S3Si	(M+3H)+3	-4.79
682.95540	682.95160	3	5079.03	C92H130N18O27S3Si	(M+3H)+3	-5.52
683.28570	683.28520	3	3240.09	C92H130N18O27S3Si	(M+3H)+3	-0.7
683.62170	683.61880	3	2090.84	C92H130N18O27S3Si	(M+3H)+3	-4.3
683.95250	683.95230	3	1496.02	C92H130N18O27S3Si	(M+3H)+3	-0.17
684.26190	684.28590	3	961.01	C92H130N18O27S3Si	(M+3H)+3	35.03
684.61090	684.61950	3	873.46	C92H130N18O27S3Si	(M+3H)+3	12.65
943.99070			17197.06			

Figure 3.93: HRMS data for peptide 3.44. Starting amount: 0.012 mmole. Yield: 5.1%.

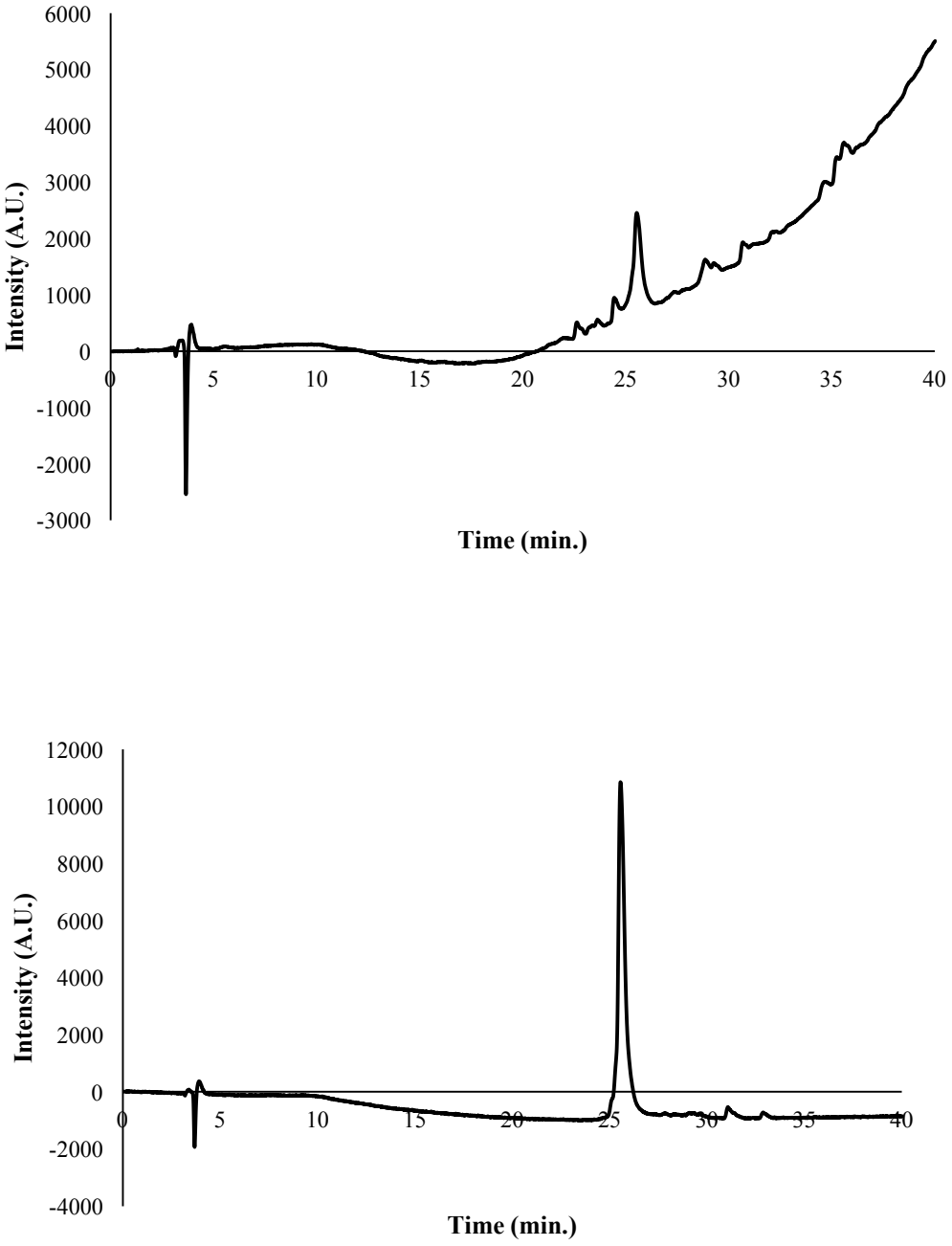
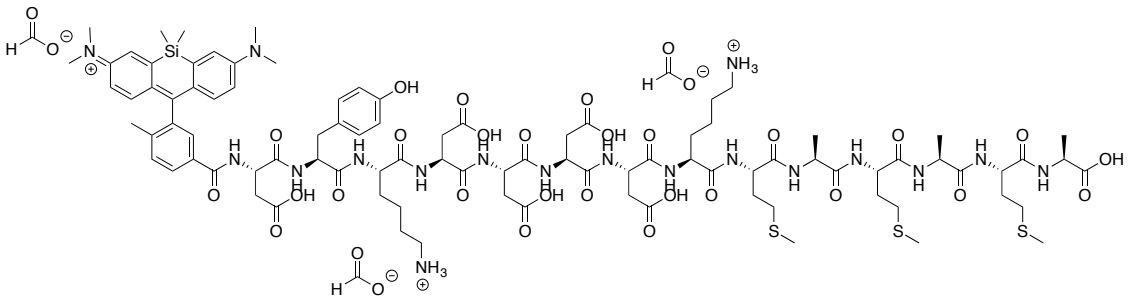
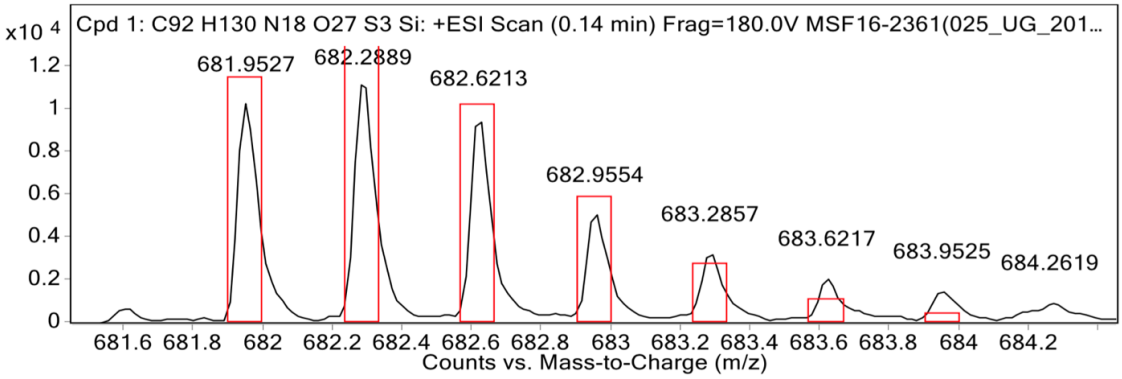


Figure 3.94: (top) Analytical trace for peptide 3.44 at 254 nm; Retention time: 25.472 min. (bottom) Analytical trace at 640 nm; Retention time: 25.51467 min.

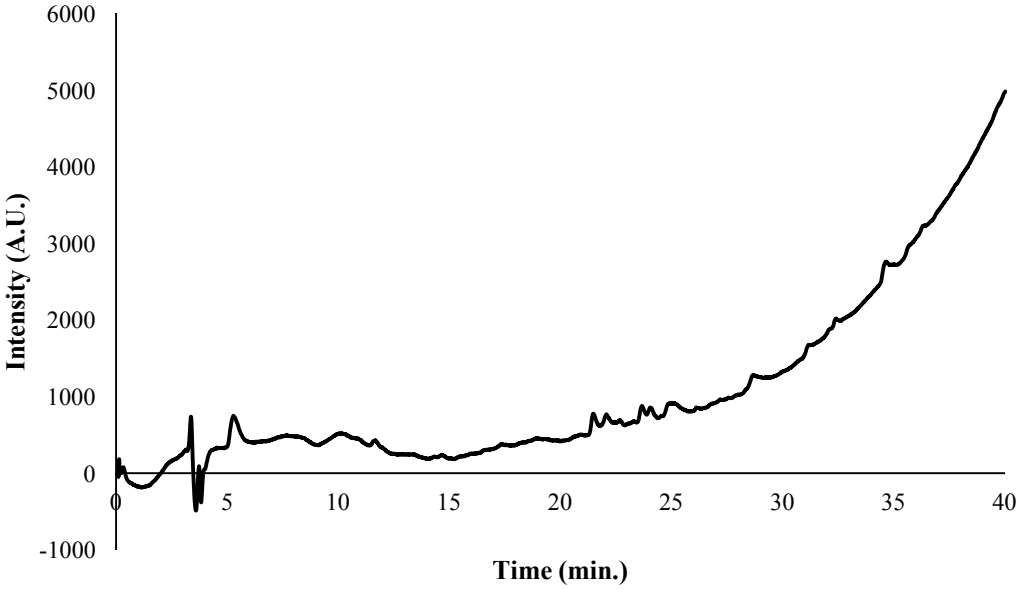




MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
681.95270	681.95010	3	10246.8	C92H130N18O27S3Si	(M+3H)+3	-3.9
682.28890	682.28430	3	11521	C92H130N18O27S3Si	(M+3H)+3	-6.71
682.62130	682.61800	3	9582.06	C92H130N18O27S3Si	(M+3H)+3	-4.79
682.95540	682.95160	3	5079.03	C92H130N18O27S3Si	(M+3H)+3	-5.52
683.28570	683.28520	3	3240.09	C92H130N18O27S3Si	(M+3H)+3	-0.7
683.62170	683.61880	3	2090.84	C92H130N18O27S3Si	(M+3H)+3	-4.3
683.95250	683.95230	3	1496.02	C92H130N18O27S3Si	(M+3H)+3	-0.17
684.26190	684.28590	3	961.01	C92H130N18O27S3Si	(M+3H)+3	35.03
684.61090	684.61950	3	873.46	C92H130N18O27S3Si	(M+3H)+3	12.65
943.99070			17197.06			

Figure 3.95: HRMS data for peptide 3.45. Starting amount: 0.012 mmole. Yield: 0.37%.



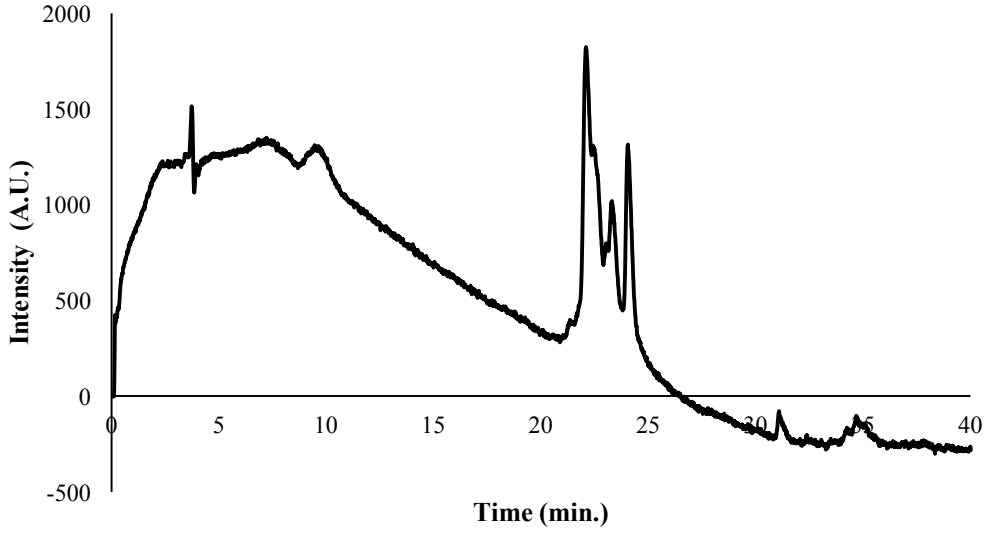
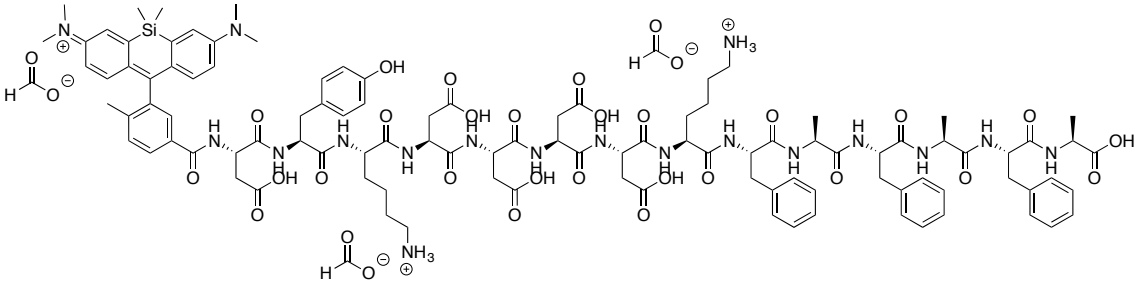
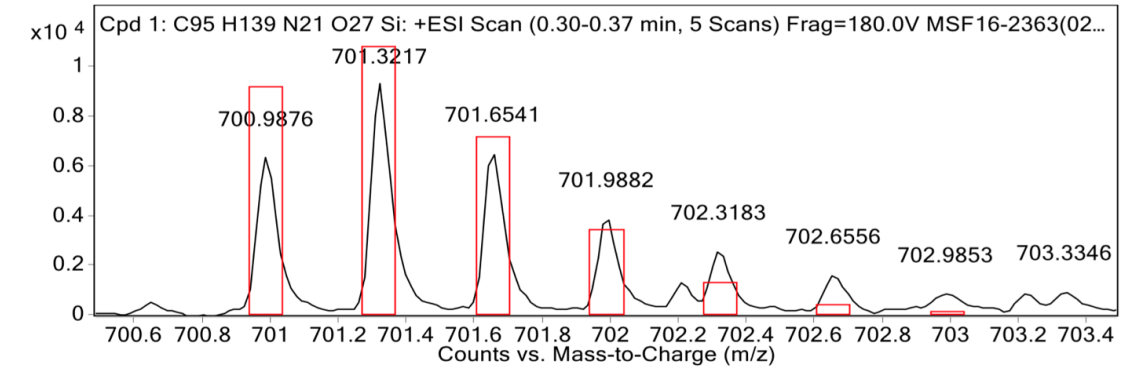


Figure 3.96: (top) Analytical trace for peptide 3.45 at 254 nm. (bottom) Analytical trace at 640 nm.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
679.34360			15318.37			
700.98760	700.98650	3	6382.74	C95H139N21O27Si	(M+3Na)+3	-1.55
701.32170	701.32080	3	9367.29	C95H139N21O27Si	(M+3Na)+3	-1.34
701.65410	701.65490	3	6563.82	C95H139N21O27Si	(M+3Na)+3	1.18
701.98820	701.98900	3	3919.56	C95H139N21O27Si	(M+3Na)+3	1.1
702.31830	702.32300	3	2602.03	C95H139N21O27Si	(M+3Na)+3	6.75
702.65560	702.65700	3	1632.5	C95H139N21O27Si	(M+3Na)+3	2.08
702.98530	702.99110	3	886.99	C95H139N21O27Si	(M+3Na)+3	8.15
703.33460	703.32510	3	974.86	C95H139N21O27Si	(M+3Na)+3	-13.53

Figure 3.97: HRMS data for peptide 3.46. Starting amount: 0.012 mmole. Yield: 2.1%.

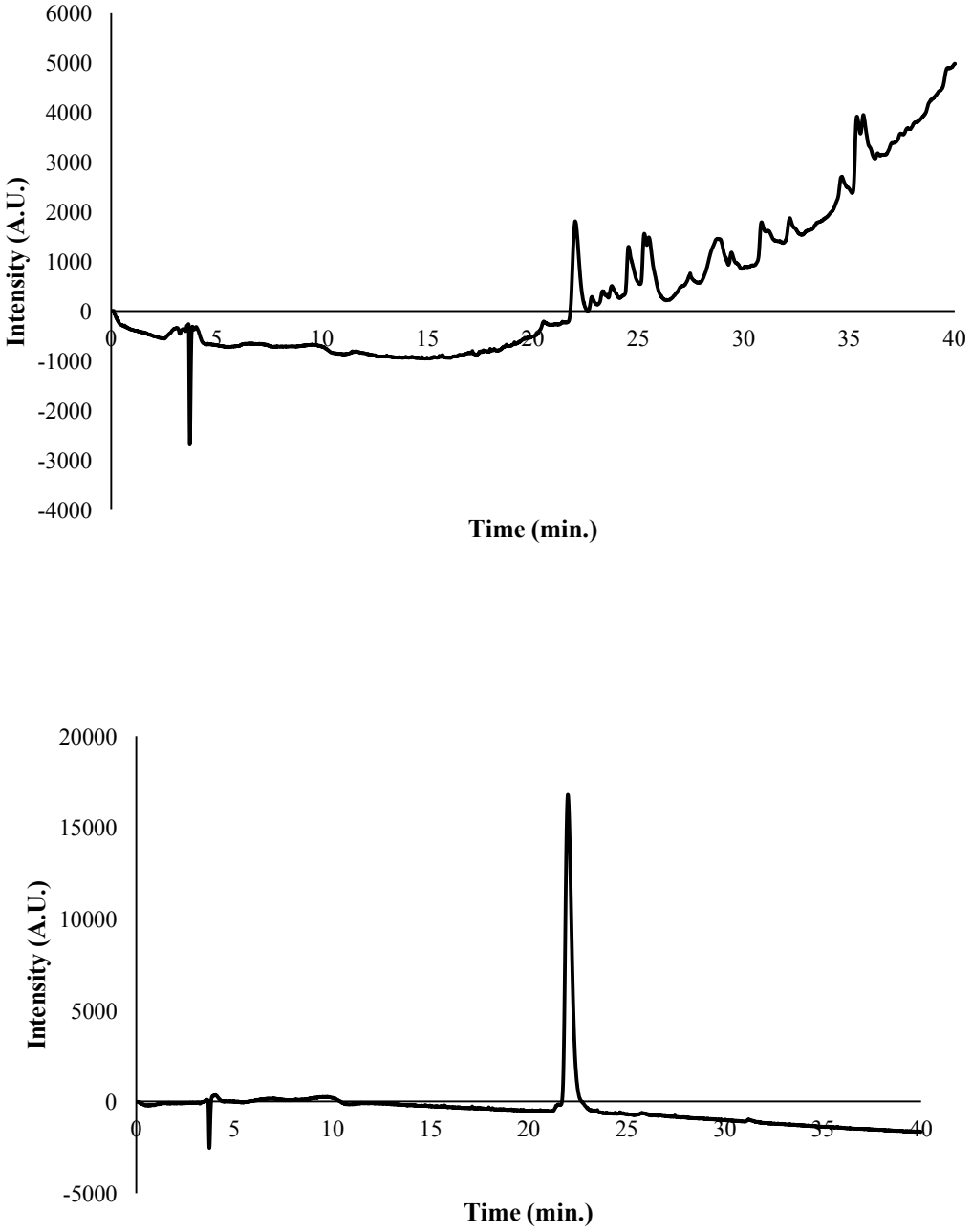
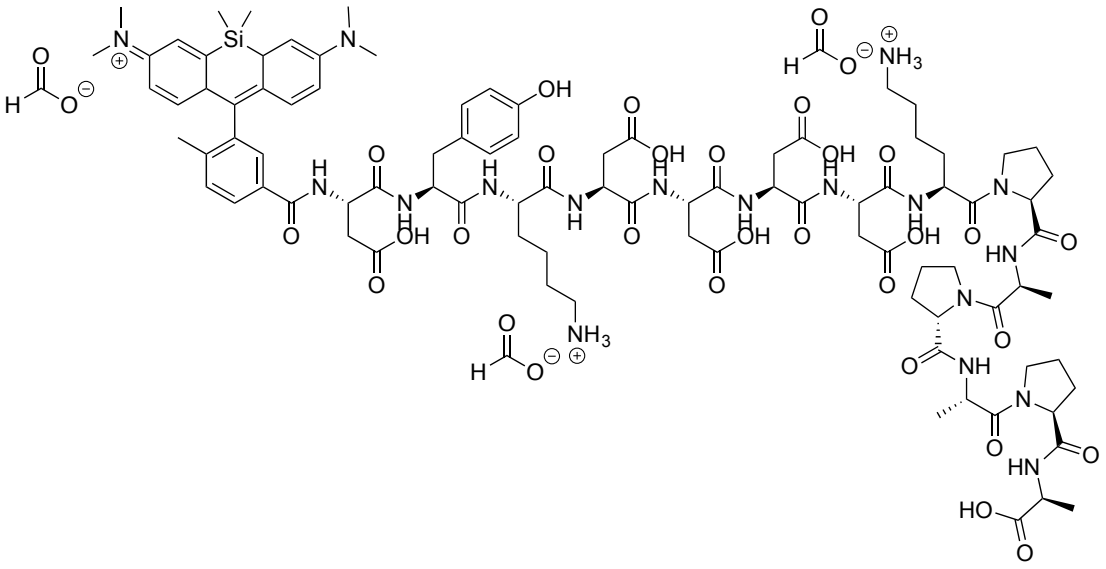
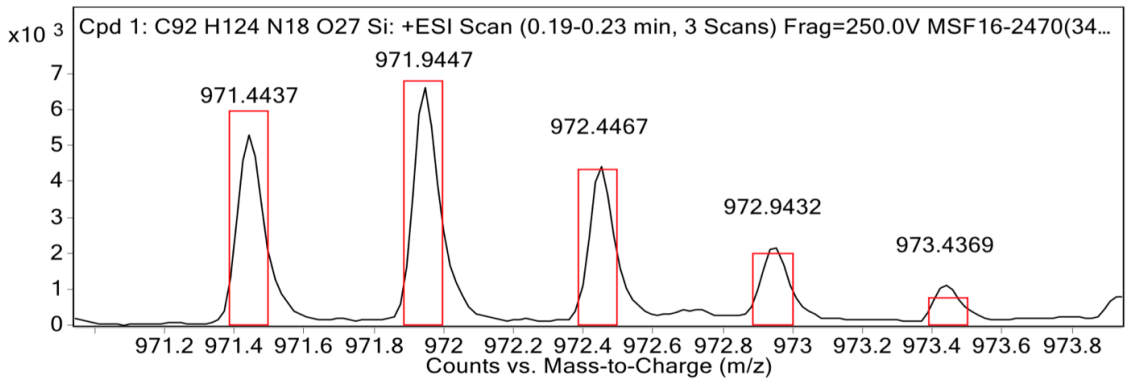


Figure 3.98: (top) Analytical trace for peptide 3.46 at 254 nm; Retention time: 21.93067 min. (bottom) Analytical trace at 640 nm; Retention time: 21.96267 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
283.12810			488803.57			
971.44370	971.43990	2	5332.4	C92H124N18O27Si	(M+2H)+2	-3.86
971.94470	971.94130	2	6662.75	C92H124N18O27Si	(M+2H)+2	-3.47
972.44670	972.44250	2	4462.87	C92H124N18O27Si	(M+2H)+2	-4.27
972.94320	972.94360	2	2223.53	C92H124N18O27Si	(M+2H)+2	0.44
973.43690	973.44470	2	1159.74	C92H124N18O27Si	(M+2H)+2	7.96

Figure 3.99: HRMS data for peptide 3.47. Starting amount: 0.012 mmole. Yield: 5.7%.

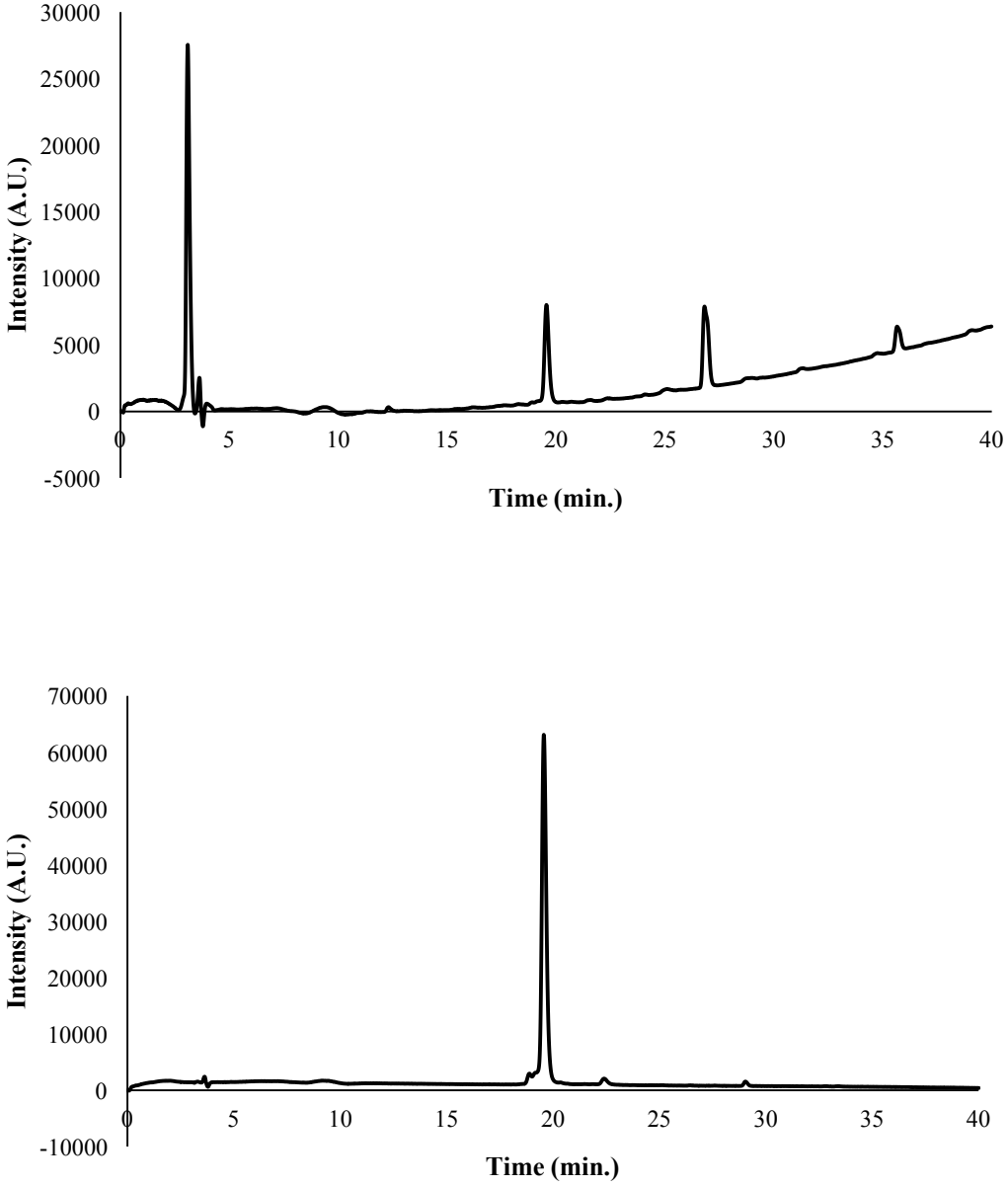
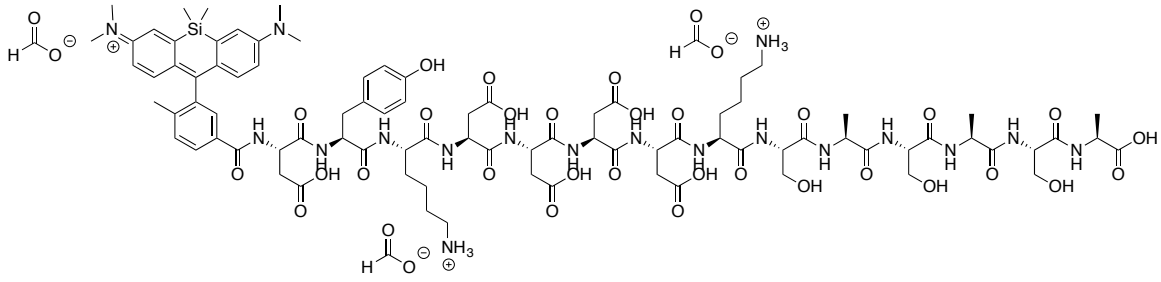
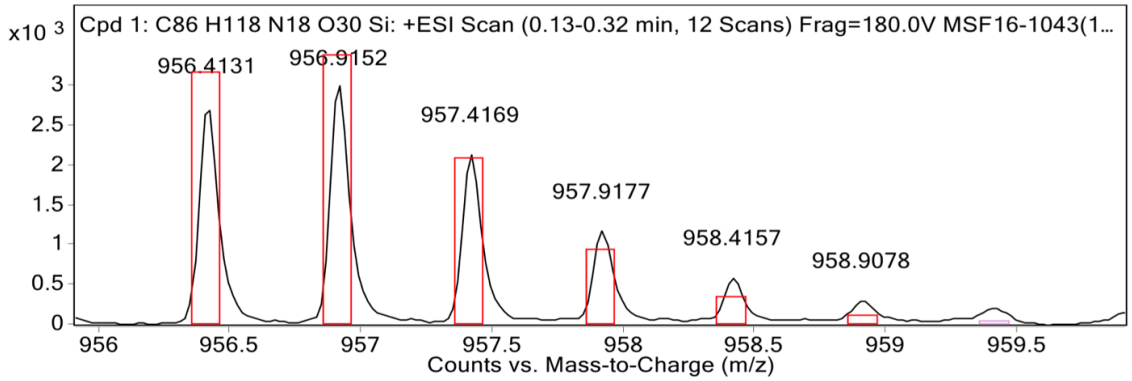


Figure 3.100: (top) Analytical trace for peptide 3.47 at 254 nm; Retention times: 3.09333 min., 19.53067 min., 26.79467 min., 35.56267 min. (bottom) Analytical trace at 640 nm; Retention time: 19.552 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
638.28100			56688.04			
956.41310	956.40880	2	2762.13	C86H118N18O30Si	(M+2H)+2	-4.53
956.91520	956.91020	2	3031.22	C86H118N18O30Si	(M+2H)+2	-5.22
957.41690	957.41140	2	2148.55	C86H118N18O30Si	(M+2H)+2	-5.77
957.91770	957.91240	2	1181.8	C86H118N18O30Si	(M+2H)+2	-5.51
958.41570	958.41340	2	597.14	C86H118N18O30Si	(M+2H)+2	-2.4
958.90780	958.91450	2	311.12	C86H118N18O30Si	(M+2H)+2	6.95

Figure 3.101: HRMS data for peptide 3.48. Starting amount: 0.012 mmole. Yield: 3.7%.

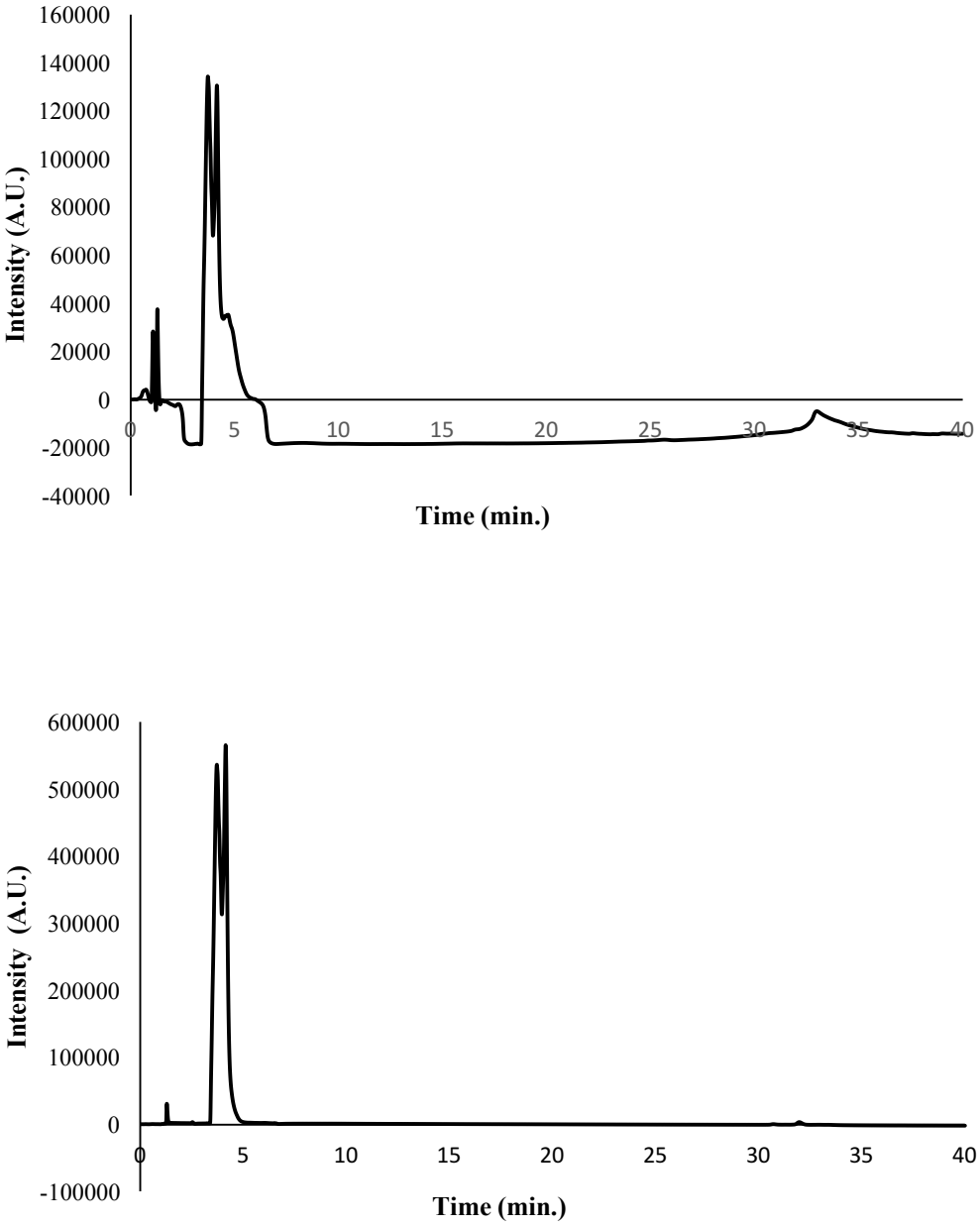
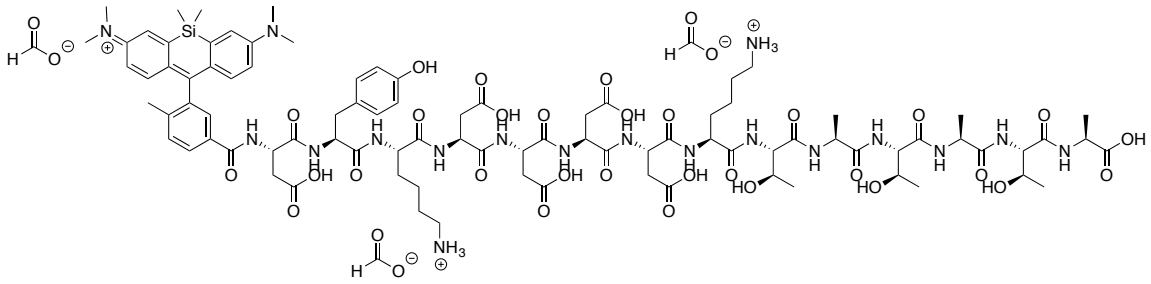
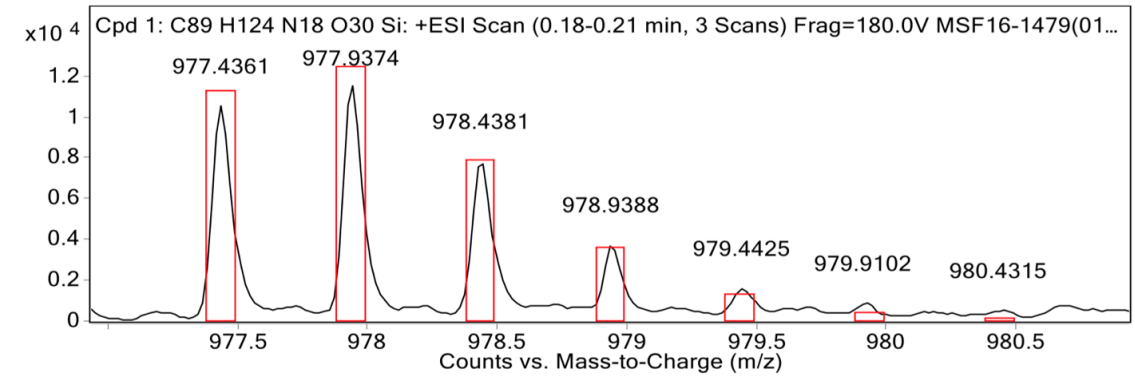


Figure 3.102: (top) Analytical trace for peptide 3.48 at 254 nm; Retention time: 3.70133 min., 3.78667 min. (bottom) Analytical trace at 640 nm; Retention time: 3.70133 min., 3.78667 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
652.29620			264735.57			
977.43610	977.43230	2	10611.02	C89H124N18O30Si	(M+2H)+2	-3.91
977.93740	977.93370	2	11630.25	C89H124N18O30Si	(M+2H)+2	-3.76
978.43810	978.43490	2	7905.64	C89H124N18O30Si	(M+2H)+2	-3.31
978.93880	978.93590	2	3813.62	C89H124N18O30Si	(M+2H)+2	-2.95
979.44250	979.43700	2	1673.8	C89H124N18O30Si	(M+2H)+2	-5.62
979.91020	979.93800	2	944.47	C89H124N18O30Si	(M+2H)+2	28.43
980.43150	980.43910	2	592.09	C89H124N18O30Si	(M+2H)+2	7.75

Figure 3.103: HRMS data for peptide 3.49. Starting amount: 0.012 mmole. Yield: 4.3%.

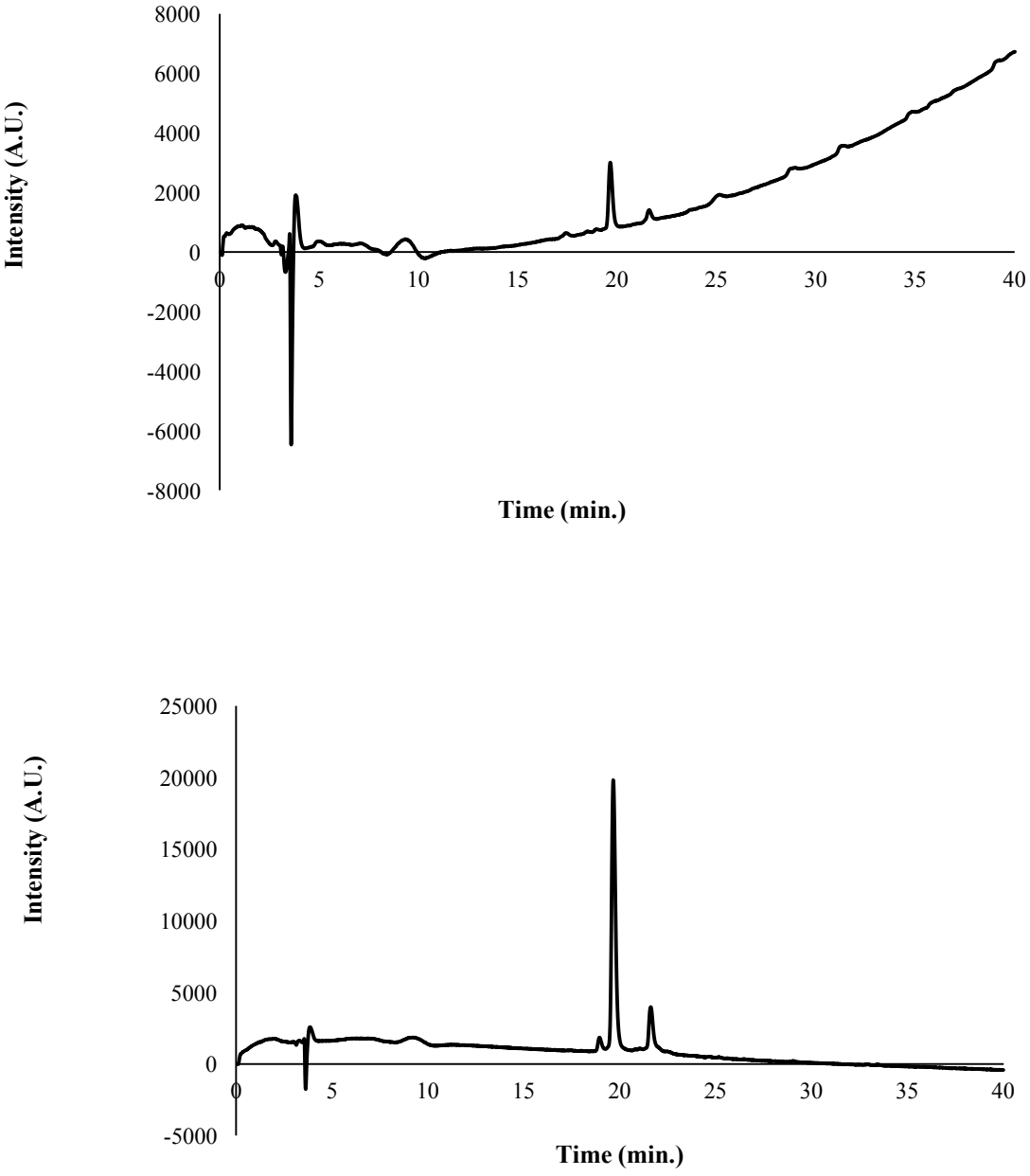
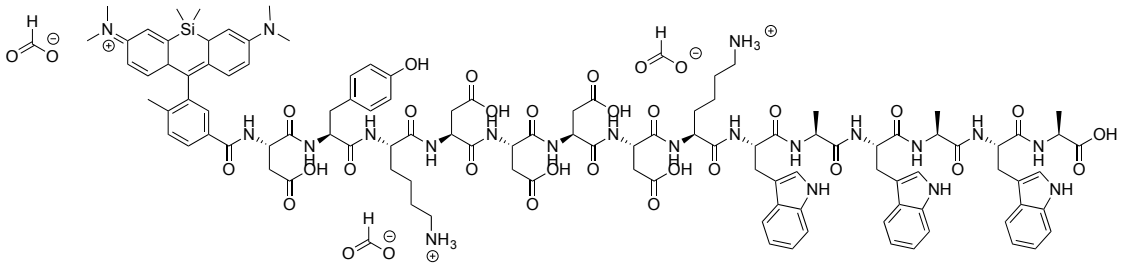
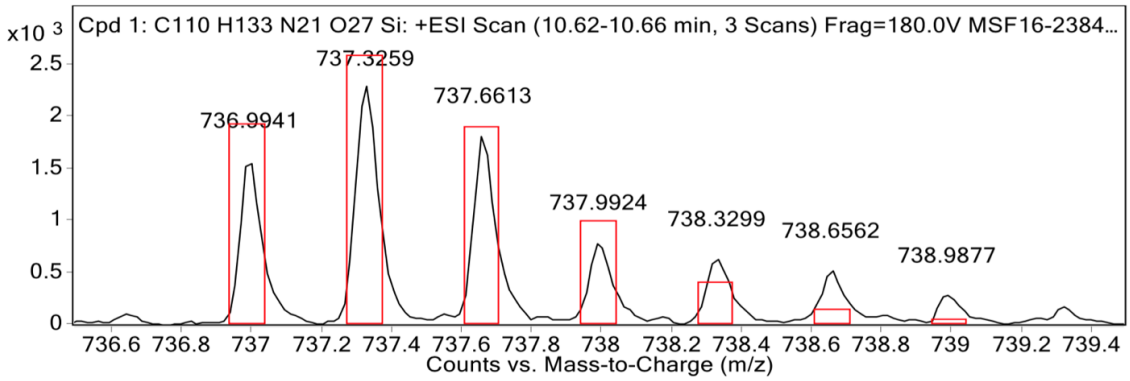


Figure 3.104: (top) Analytical trace for peptide 3.49 at 254 nm; Retention time: 19.616 min. (bottom) Analytical trace at 640 nm; Retention time: 19.648 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
736.99410	736.98890	3	1586.14	C110H133N21O27Si	(M+3H)+3	-7.1
737.32590	737.32320	3	2304.07	C110H133N21O27Si	(M+3H)+3	-3.69
737.66130	737.65740	3	1825.5	C110H133N21O27Si	(M+3H)+3	-5.38
737.99240	737.99150	3	796.61	C110H133N21O27Si	(M+3H)+3	-1.26
738.32990	738.32560	3	637.44	C110H133N21O27Si	(M+3H)+3	-5.82
738.65620	738.65960	3	524.99	C110H133N21O27Si	(M+3H)+3	4.6
738.98770	738.99370	3	287.12	C110H133N21O27Si	(M+3H)+3	8.11
943.97870			4760.09			

Figure 3.105: HRMS data for peptide 3.50. Starting amount: 0.012 mmole. Yield: 2.7%.

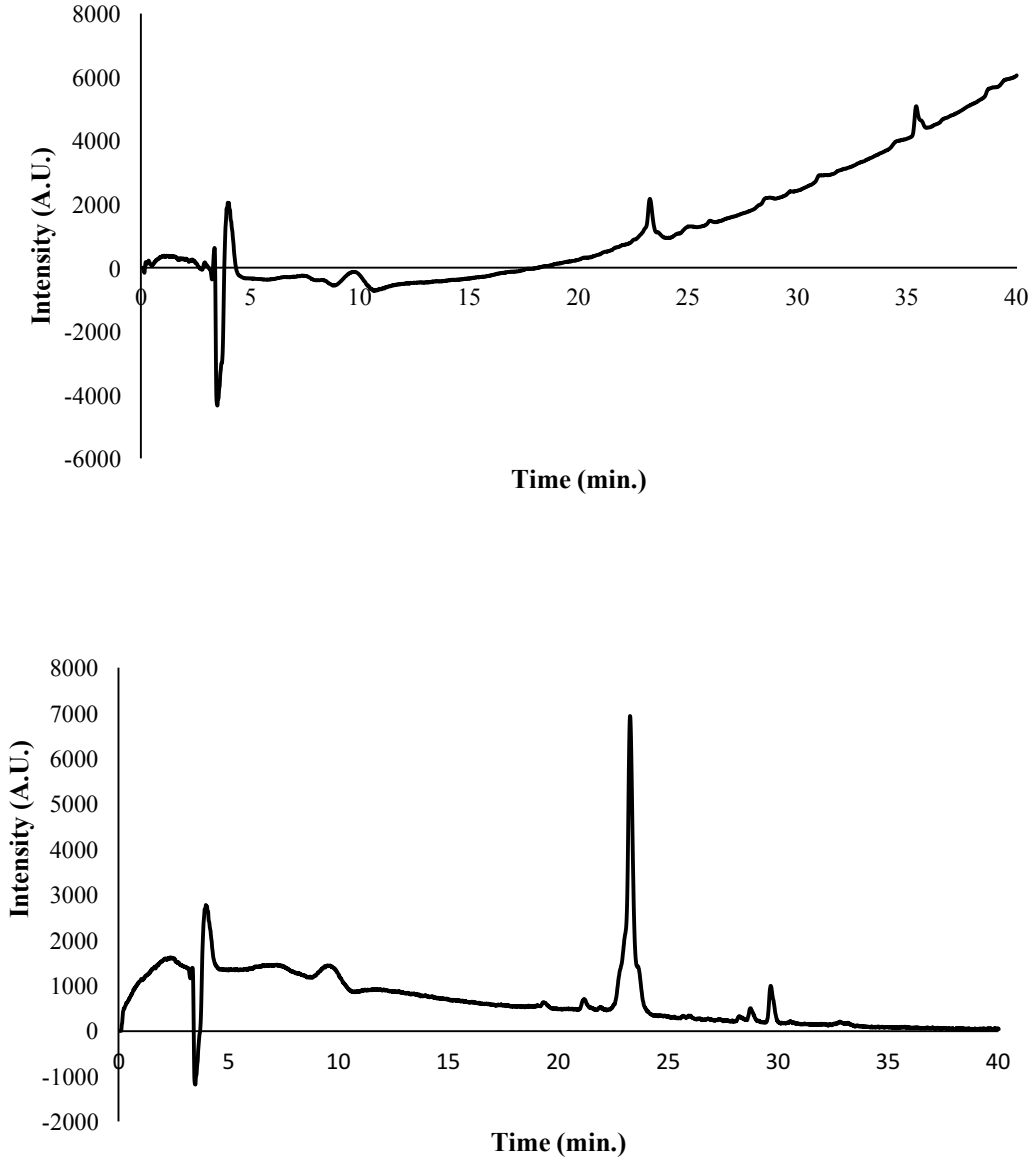
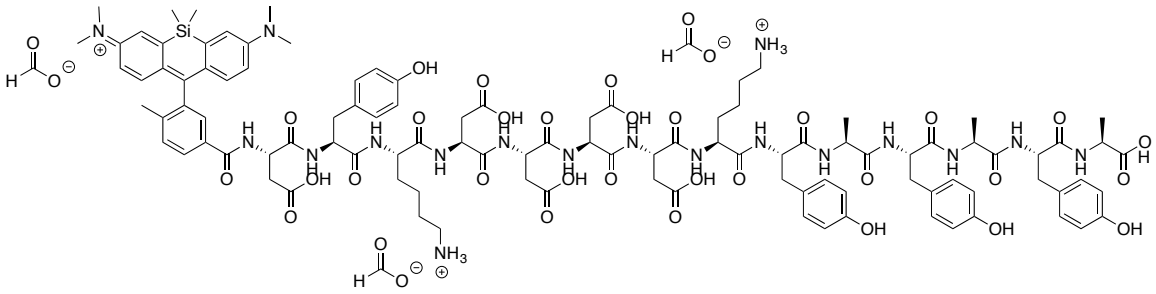
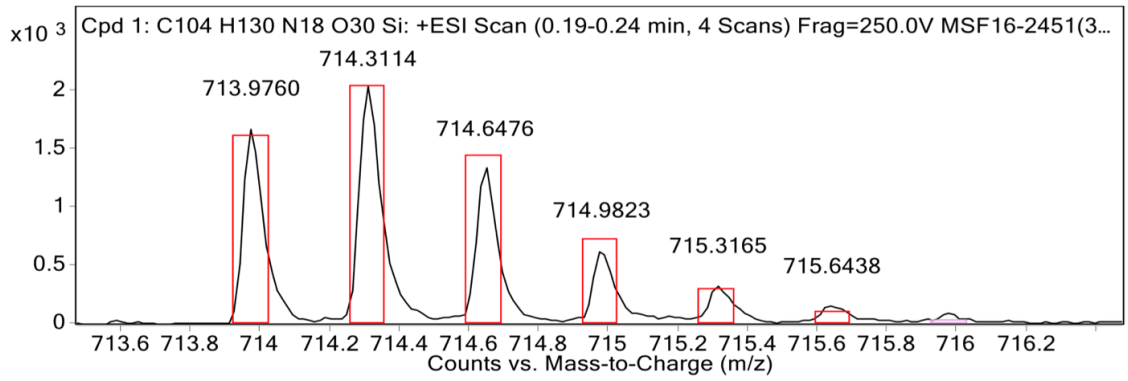


Figure 3.106: (top) Analytical trace for peptide 3.51 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 23.232 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
713.97600	713.97290	3	1681.73	C104H130N18O30Si	(M+3H)+3	-4.29
714.31140	714.30720	3	2033.25	C104H130N18O30Si	(M+3H)+3	-5.79
714.64760	714.64140	3	1336.11	C104H130N18O30Si	(M+3H)+3	-8.64
714.98230	714.97550	3	636.57	C104H130N18O30Si	(M+3H)+3	-9.48
715.31650	715.30960	3	333.25	C104H130N18O30Si	(M+3H)+3	-9.64
715.64380	715.64360	3	159.57	C104H130N18O30Si	(M+3H)+3	-0.19
959.96510			18352.02			

Figure 3.107: HRMS data for peptide 3.52. Starting amount: 0.012 mmole. Yield: 1.2%.

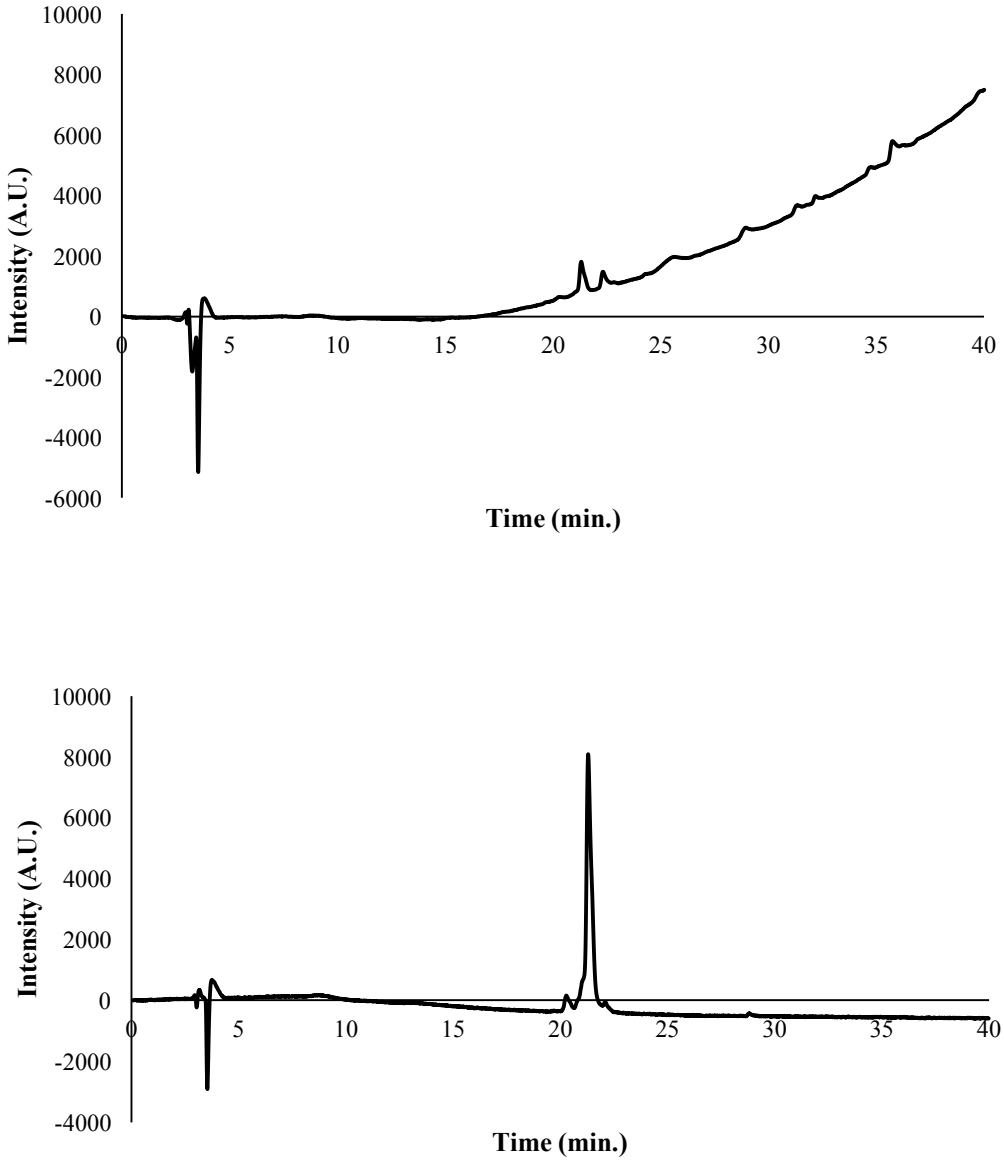
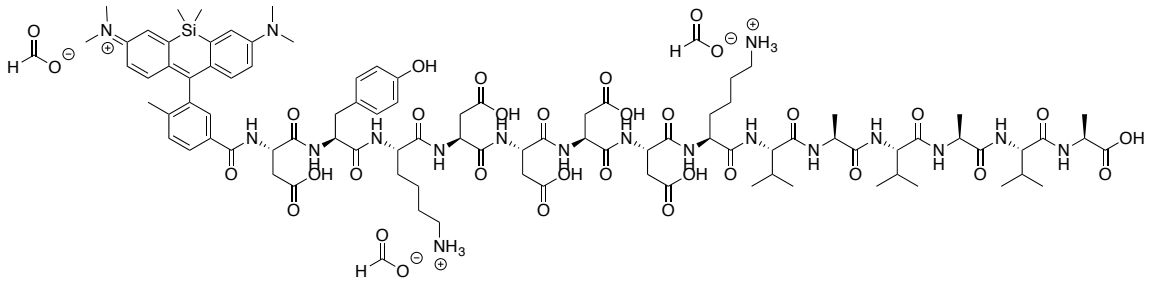
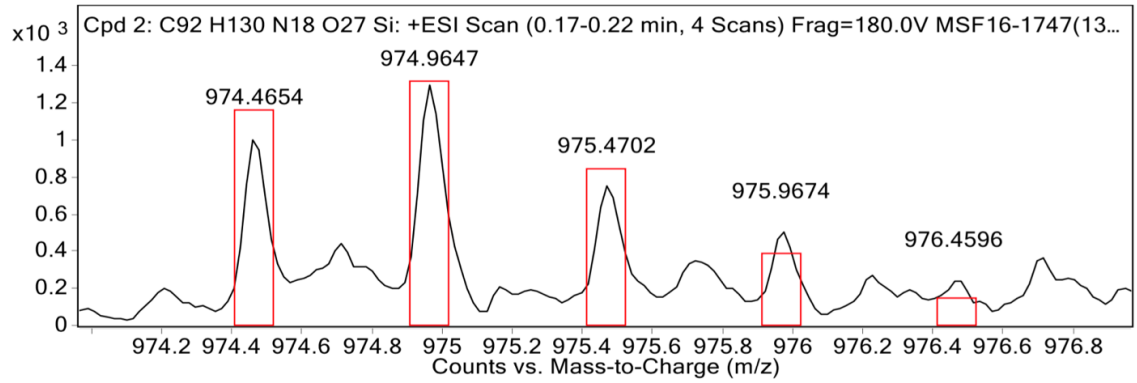


Figure 3.108: (top) Analytical trace for peptide 3.52 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 21.28 min.

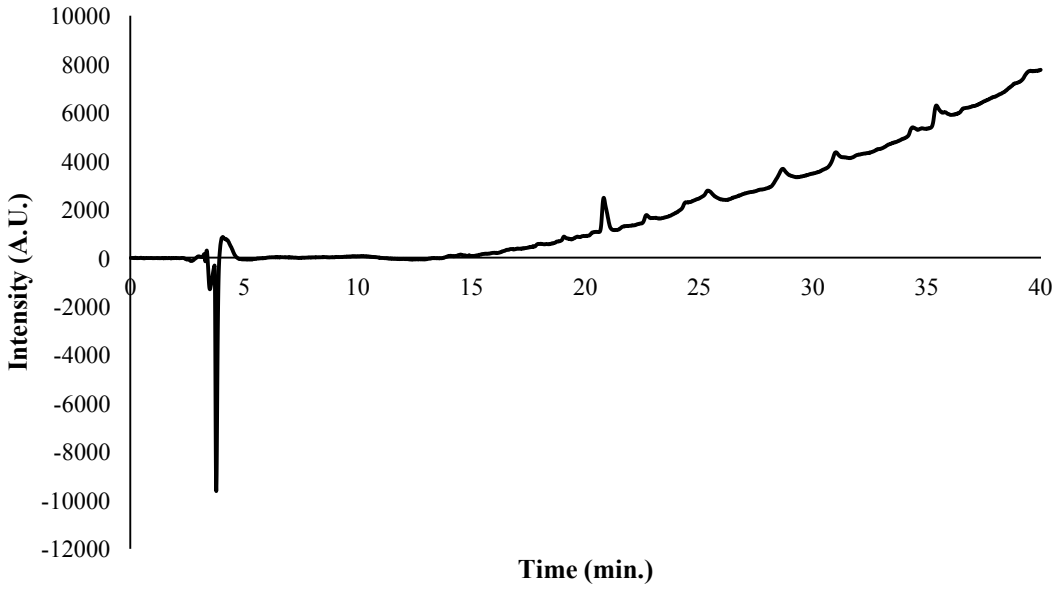




MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
959.96610			35086.35			
974.46540	974.46340	2	1025.26	C92H130N18O27Si	(M+2H)+2	-2.1
974.96470	974.96480	2	1299.06	C92H130N18O27Si	(M+2H)+2	0.12
975.47020	975.46600	2	763.33	C92H130N18O27Si	(M+2H)+2	-4.33
975.96740	975.96710	2	517.25	C92H130N18O27Si	(M+2H)+2	-0.37
976.45960	976.46810	2	254.54	C92H130N18O27Si	(M+2H)+2	8.76

Figure 3.109: HRMS data for peptide 3.53. Starting amount: 0.012 mmole. Yield: 4.4%.



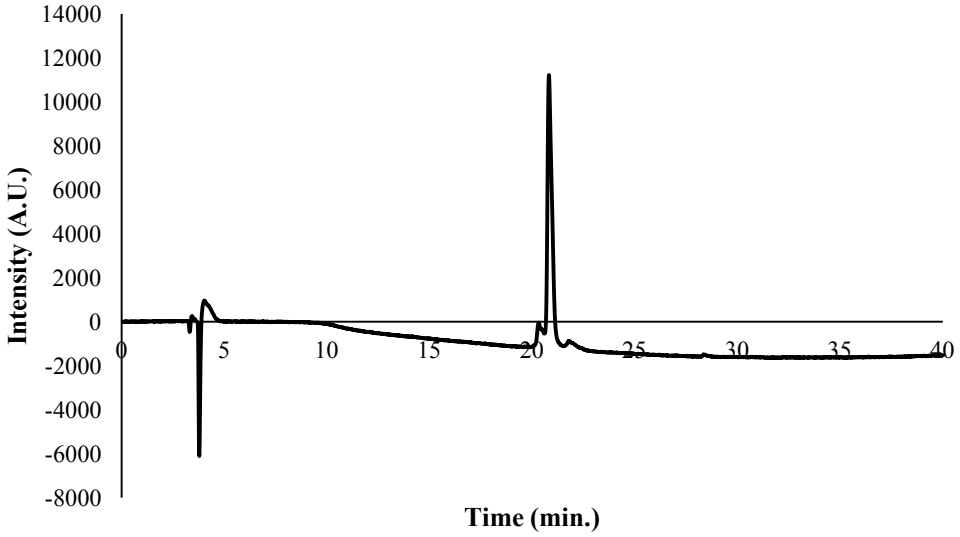
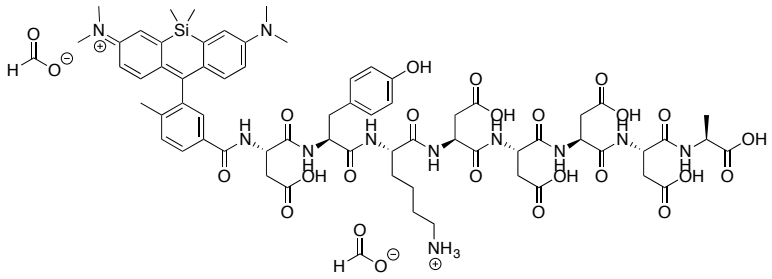
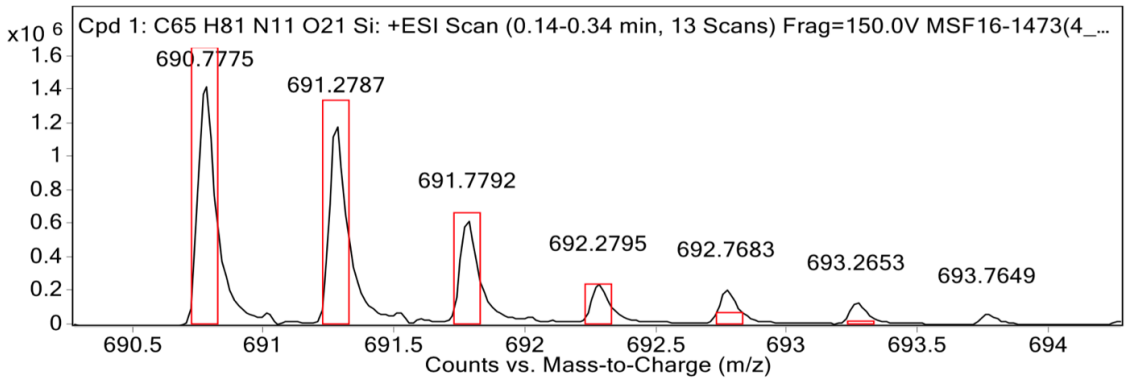


Figure 3.110: (top) Analytical trace for peptide 3.53 at 254 nm. (bottom) Analytical trace at 640 nm; Retention time: 20.78933 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
690.77750	690.77620	2	1444826.99	C65H81N11O21Si	(M+2H)+2	-1.9
691.27870	691.27760	2	1195697.21	C65H81N11O21Si	(M+2H)+2	-1.68
691.77920	691.77860	2	626011.71	C65H81N11O21Si	(M+2H)+2	-0.97
692.27950	692.27950	2	250240.87	C65H81N11O21Si	(M+2H)+2	0.02
692.76830	692.78050	2	214531.97	C65H81N11O21Si	(M+2H)+2	17.58
693.26530	693.28150	2	136487.46	C65H81N11O21Si	(M+2H)+2	23.36
693.76490	693.78250	2	59269.56	C65H81N11O21Si	(M+2H)+2	25.42
694.26570	694.28360	2	22097.66	C65H81N11O21Si	(M+2H)+2	25.71
694.75790	694.78460	2	23090.44	C65H81N11O21Si	(M+2H)+2	38.44

Figure 3.111: HRMS data for peptide 3.54. Starting amount: 0.012 mmole. Yield: 0.94%.

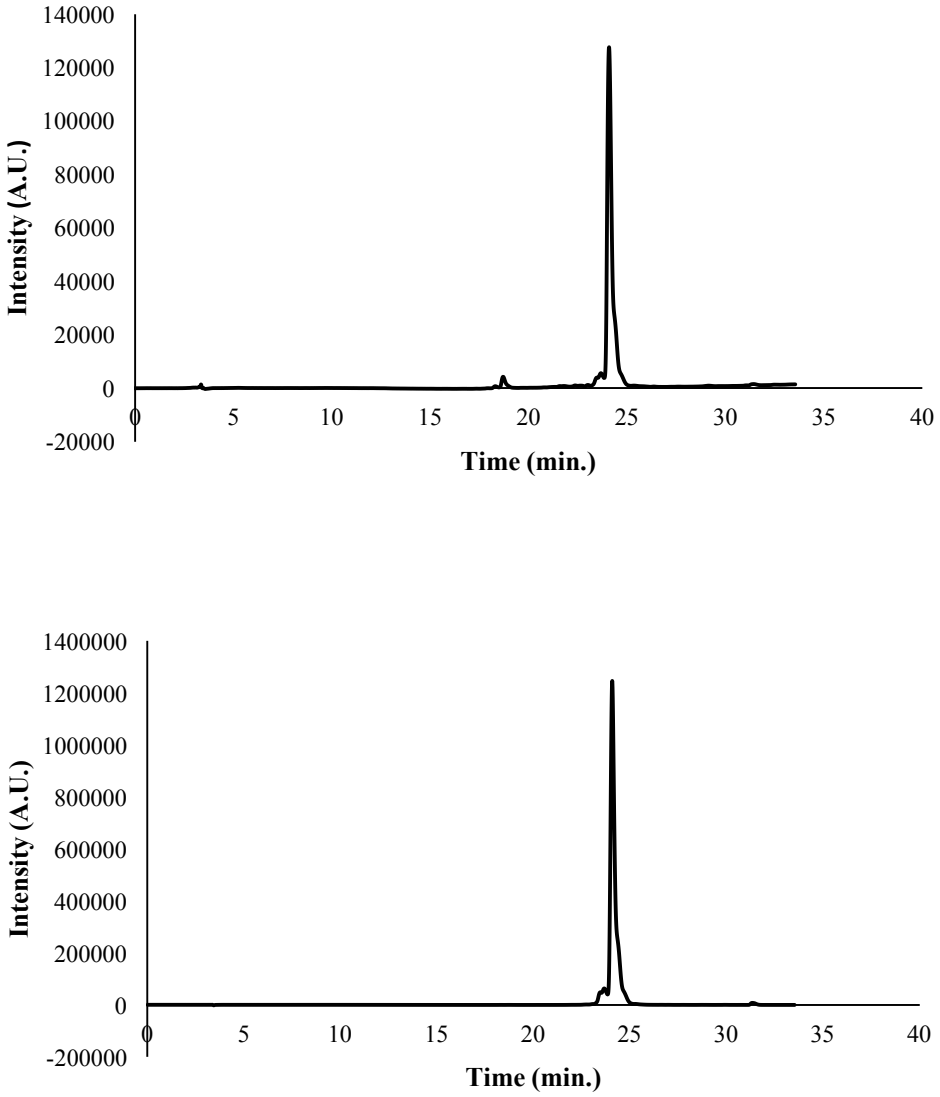
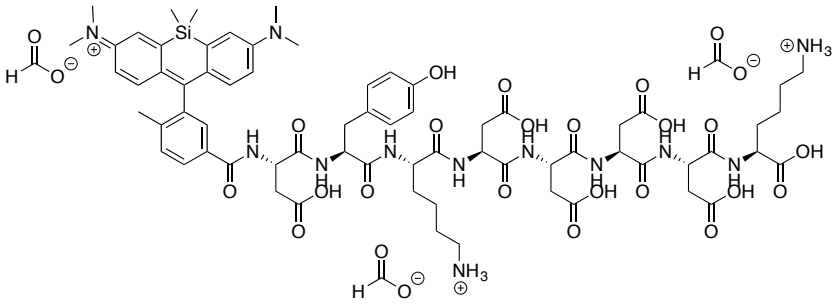
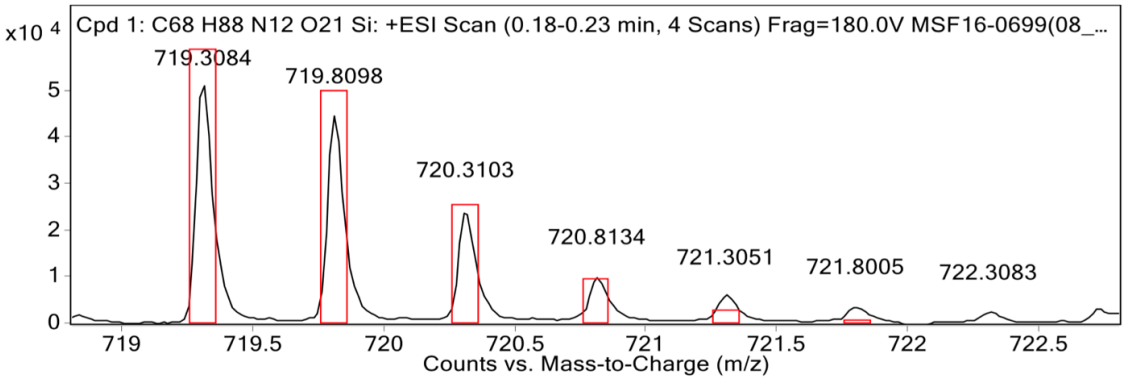


Figure 3.112: (top) Analytical trace for peptide 3.54 at 254 nm; Retention time: 24.05333 min. (bottom) Analytical trace at 640 nm; Retention time: 24.05333 min.





MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
719.30840	719.30510	2	51818.99	C68H88N12O21Si	(M+2H)+2	-4.59
719.80980	719.80650	2	44780.39	C68H88N12O21Si	(M+2H)+2	-4.58
720.31030	720.30750	2	24649.48	C68H88N12O21Si	(M+2H)+2	-3.88
720.81340	720.80850	2	10233.85	C68H88N12O21Si	(M+2H)+2	-6.79
721.30510	721.30950	2	5894.84	C68H88N12O21Si	(M+2H)+2	6.1
721.80050	721.81050	2	3840.63	C68H88N12O21Si	(M+2H)+2	13.79
722.30830	722.31150	2	2653.02	C68H88N12O21Si	(M+2H)+2	4.42
722.75670	722.81250	2	3383.57	C68H88N12O21Si	(M+2H)+2	77.27
763.36130			264470.77			

Figure 3.113: HRMS data for peptide 3.55. Starting amount: 0.012 mmole. Yield: 0.66%.

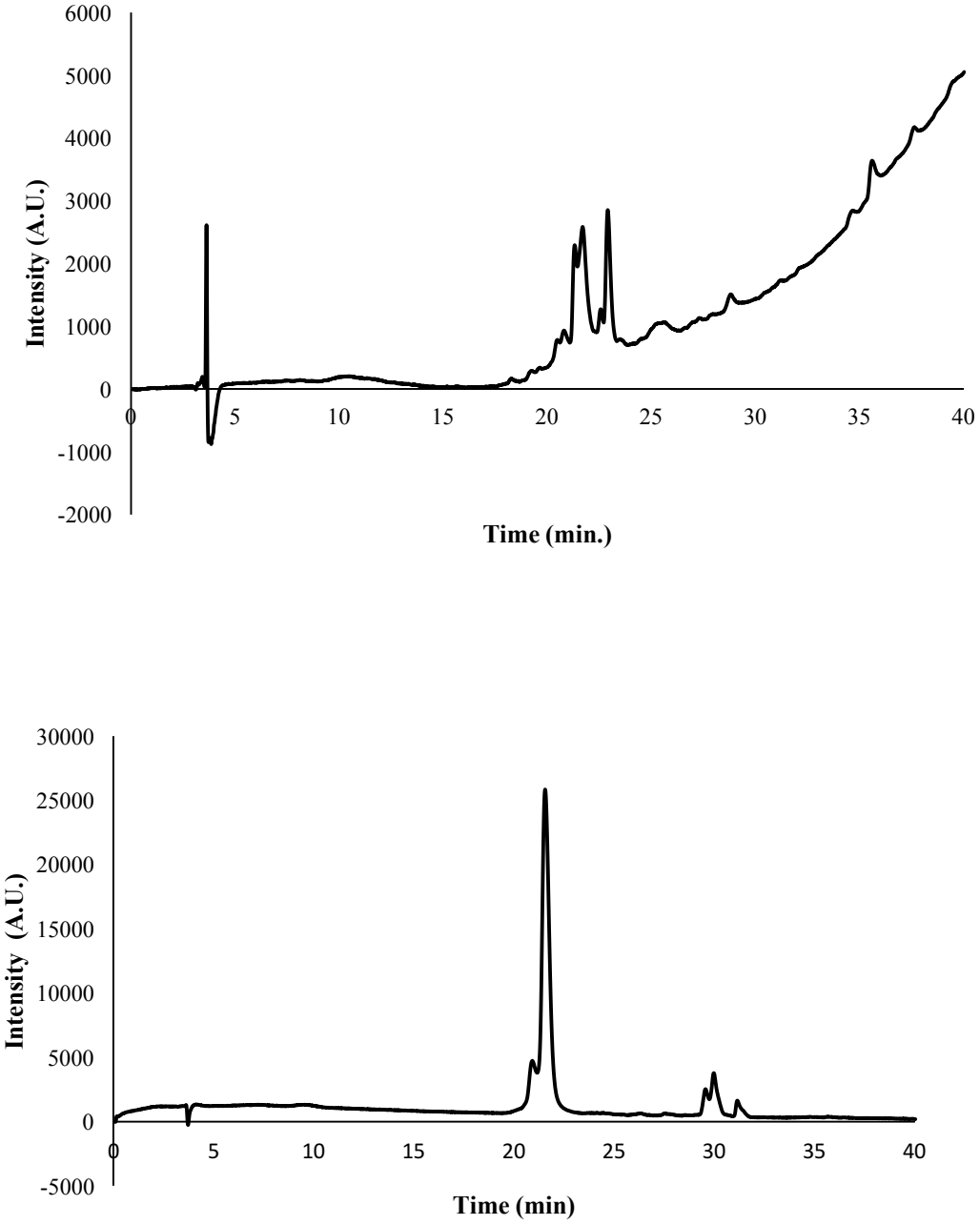


Figure 3.114: (top) Analytical trace for peptide 3.55 at 254 nm; Retention times: 21.26933 min., 22.85867 min. (bottom) Analytical trace at 640 nm; Retention time: 21.504 min.

3.7 References

1. M. Larance and A. I. Lamond, *Nat. Rev. Mol. Cell. Biol.*, 2015, **16**, 269-280.
2. Y. V. Karpievitch, A. D. Polpitiya, G. A. Anderson, R. D. Smith and A. R. Dabney, *Ann. Appl. Stat.*, 2010, DOI: 10.1214/10-AOAS341, 1797-1823.
3. N. Schupf, M. X. Tang, H. Fukuyama, J. Manly, H. Andrews, P. Mehta, J. Ravetch and R. Mayeux, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 14052-14057.
4. D. Theodorescu, E. Schiffer, H. W. Bauer, F. Douwes, F. Eichhorn, R. Polley, T. Schmidt, W. Schöfer, P. Zürbig, D. M. Good, J. J. Coon and H. Mischak, *Proteomics Clin. Appl.*, 2008, **2**, 556-570.
5. K. Petritis, L. J. Kangas, B. Yan, M. E. Monroe, E. F. Strittmatter, W.-J. Qian, J. N. Adkins, R. J. Moore, Y. Xu, M. S. Lipton, D. G. Camp and R. D. Smith, *Anal. Chem.*, 2006, **78**, 5026-5039.
6. J. S. Page, R. T. Kelly, K. Tang and R. D. Smith, *J. Am. Soc. Mass. Spectrom.*, 2007, **18**, 1582-1590.
7. W.-C. Yang, H. Mirzaei, X. Liu and F. E. Regnier, *Anal. Chem.*, 2006, **78**, 4702-4708.
8. Y. Gao and Y. Wang, *J. Am. Mass. Spec.*, 2007, **18**, 1973-1976.
9. J. Swaminathan, A. A. Boulgakov and E. M. Marcotte, *PLOS Computational Biology*, 2015, **11**, e1004080.
10. S. Goodwin, J. D. McPherson and W. R. McCombie, *Nat. Rev. Genet.*, 2016, **17**, 333-351.
11. J. M. Heather and B. Chain, *Genomics*, 2016, **107**, 1-8.
12. I. Braslavsky, B. Hebert, E. Kartalov and S. R. Quake, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 3960-3964.
13. J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. deWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach and S. Turner, *Science*, 2009, **323**, 133-138.
14. S. Uemura, C. E. Aitken, J. Korlach, B. A. Flusberg, S. W. Turner and J. D. Puglisi, *Nature*, 2010, **464**, 1012-1017.
15. F. Ding, M. Manosas, M. M. Spiering, S. J. Benkovic, D. Bensimon, J.-F. Allemand and V. Croquette, *Nat. Meth,*, 2012, **9**, 367-372.
16. N. Bandeira, V. Pham, P. Pevzner, D. Arnott and J. R. Lill, *Nat. Biotech.*, 2008, **26**, 1336-1338.
17. Y. Yao, D. Margreet, G. Jetty van, R. Dick de and J. Chirlmin, *Phys. Biol.*, 2015, **12**, 055003.
18. T. Karstens and K. Kobs, *J. Phys. Chem.*, 1980, **84**, 1871-1872.
19. Y. A. Andreev, S. A. Kozlov, A. A. Vassilevski and E. V. Grishin, *Anal. Biochem.*, 2010, **407**, 144-146.

20. A. N. Zaykov, J. P. Mayer and R. D. DiMarchi, *Nat. Rev. Drug Discov.*, 2016, **15**, 425-439.
21. D. L. Swaney, C. D. Wenger and J. J. Coon, *J. Proteome Res.*, 2010, **9**, 1323-1329.
22. Y. Koide, Y. Urano, K. Hanaoka, T. Terai and T. Nagano, *J. Am. Chem. Soc.*, 2011, **133**, 5680-5682.
23. T. P. Hopp, K. S. Prickett, V. L. Price, R. T. Libby, C. J. March, D. Pat Cerretti, D. L. Urdal and P. J. Conlon, *Nat. Biotech.*, 1988, **6**, 1204-1210.

Chapter 4: An Efficient Methodology to Introduce *o*-(aminomethyl)phenyl-boronic Acids into Peptides: Alkylation of Secondary Amines[4]

4.1 INTRODUCTION

4.1.1 Need for Facile Incorporation of *o*-(aminomethyl)phenyl-boronic Acids into Peptides

Boronic acid containing compounds continue to gain interest in the areas of chemosensing, materials chemistry, fluorescence imaging, mass spectrometry, and biomedical engineering.[1-22] From the sensing of sugars to the formation of dynamic covalent structures, or serving as probes for the enrichment of analytes, boronic acid compounds are seen as an important organic moiety for developing novel analytical and *in vivo* applications. Continued growth in the biological sciences has sparked interest for synthesizing boronic acid containing peptides. These are attractive synthetic targets because peptides are versatile platforms for incorporating binding specificity to a target of interest. Boronic acid peptides could also be used in the design of combinatorial libraries. A method for incorporating a boronic acid into the side chains at controllable positions increases the diversity of compounds accessible.

4.1.2 Current Approaches to Incorporate Boronic Acids into Peptides

Common approaches for synthesizing boronic acid containing peptides include the use of building blocks with a boronate ester side chain, amidation of boronic acid compounds, or reductive amination of residues with 2-formylphenyl boronic acid.[23-31] The most common approach is incorporating a building block during solid-phase synthesis.

---

4 **Hernandez, Erik**; Kolesnichenko, Igor; Reuther, James F. "Secondary-Amine Introduction to Peptides and Subsequent Derivatization with 2-(Bromomethyl)phenylboronic acid." *New Journal of Chemistry*. **2017**, *41*, 126-133. Designed and synthesized peptides demonstrating chemistries to append boronic acids.

However, long synthetic routes to non-commercially available boronic acid reactants are required (Figure 4.1a). To derivatize the C-terminus, amidation of amino boronic acids is usually used (Figure 4.1b). Peptides with aspartates or glutamates have to be avoided. If carboxylate side chains are used, no site- specific reactivity can be achieved. Such a limitation decreases the diversity of peptides available. Lastly, reductive amination, performed previously in our group, consisted of modifying only one amine residue (Figure 4.1c).

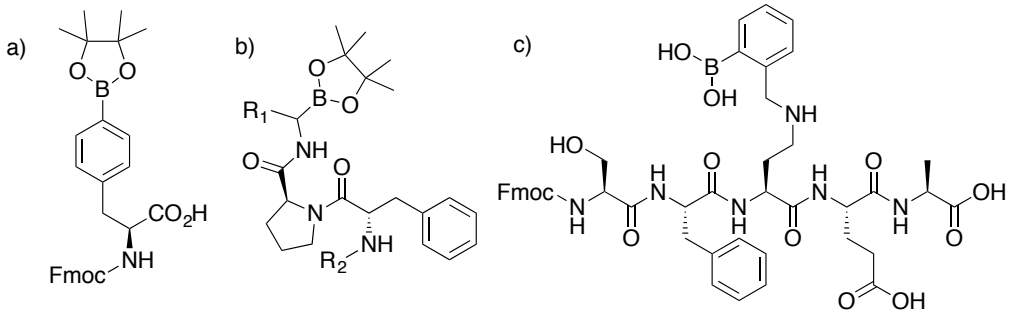


Figure 4.1: Boronic Acids. a) Synthesized building block. Requires a total of five synthetic steps. b) Boronate ester moieties introduced at the C-terminus via coupling conditions. c) Polypeptide with ortho-aminomethyl boronic acid, synthesized by reductive amination of 2-formylphenyl boronic acid.

4.1.3 Limitations with Current Approaches

Most boronic acids do not bind saccharides efficiently at neutral pH. While electron withdrawing groups on the phenyl ring, such as in *m*-nitrophenylboronic acid, can act to promote boronate ester formation at neutral pH, the most common structural feature that promotes binding is an *o*-aminomethyl group.[32, 33] This structural feature is often overlooked when incorporating a boronic acid into a receptor meant to be used in a biological setting. The only previously published strategy that allows for easy incorporation of an *o*-aminomethyl group is reductive amination. Alternatively, it would

be feasible to incorporate the *o*-aminomethyl moiety into an amino acid building block in advance of peptide synthesis, but this would further increase the number of synthetic steps. The methods we present in this paper were devised to be synthetically simple and afford the critical *o*-aminomethyl group.

4.1.4 Scientific Aim: Use Alkylation Strategies and Secondary Amines to Incorporate *o*-(aminomethyl)phenyl-boronic Acids

To increase the number of targetable sites on peptides while minimizing the difficulty and time required to synthesize *o*-(aminomethyl)phenylboronic acid building blocks, we developed chemistries for functionalizing peptides with secondary amines. Primary amines are present in the side chains of natural amino acids such as lysine. However, alkylation of primary amines typically suffers from over-alkylation. If secondary amines residues are present, then alkylation approaches become a viable alternative. The commercial availability of *o*-(bromomethyl) phenylboronic acid is an additional benefit to an alkylation approach.

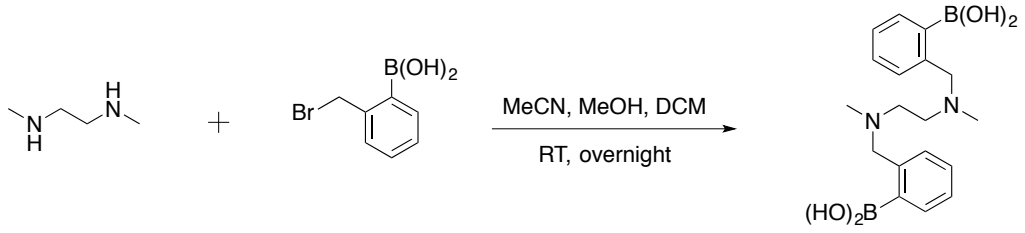
Herein, we report three methods that introduce secondary amines and subsequent *o*-(aminomethyl) phenylboronic acids to peptides. The methods include both solid-phase and solution-phase approaches. The first approach was the selective modification of cysteine residues in solution using an iodoacetamide linker with a boc-protected secondary amine. In the second approach, a cysteamine resin and an azido lysine building block was used in the solid-phase synthesis of peptides with a secondary amine. The third approach, using the commercially available N_{ε}-methyl lysine, was the most direct way for making our desired peptides. Once a secondary amine was incorporated, a reaction with *o*-(bromomethyl)phenylboronic acid afforded the *o*-aminomethyl moiety on all our peptides.

Finally, to demonstrate the usefulness of our chemistry for making complex structures, seven peptides containing N_{ε}-methyl lysine amino acids and derivatized at the N-terminus with a pegylated biotin were synthesized.

4.2. Strategies to Incorporate Secondary Amines into Peptides

4.2.1 Solution-Phase Incorporation of Secondary Amines and Alkylation

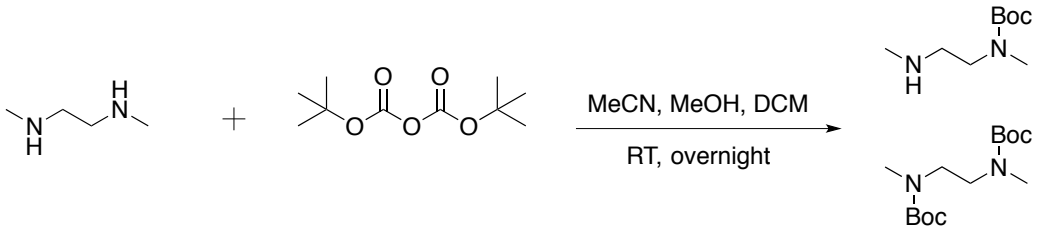
To first test if an alkylation strategy would be successful, N, N'-dimethylethylenediamine (DMEDA) was used as a model (Scheme 4.1). The alkylation occurred with a purified yield of 16% (**1**). As is well recognized by the boronic acid community[34,] boronic acids are notoriously difficult to isolate in pure form due to streaking on almost all chromatographic methods. Due to this, pinacol-protected boronate esters are more commonly used due to ease of handling and isolation of boronic acid compounds. However, the reaction conditions needed for deprotection of the boronic ester were considered to be too harsh for peptides, and thus unprotected boronic acids were chosen. Irrespective of the low yield of this model reaction we pushed forward with peptide derivatization because of the ease of the technique. Mass spectrometry and ^{1}H- NMR confirmed the presence of the desired product. The mass spectra of this product and many others in this report exhibited a characteristic loss of water.



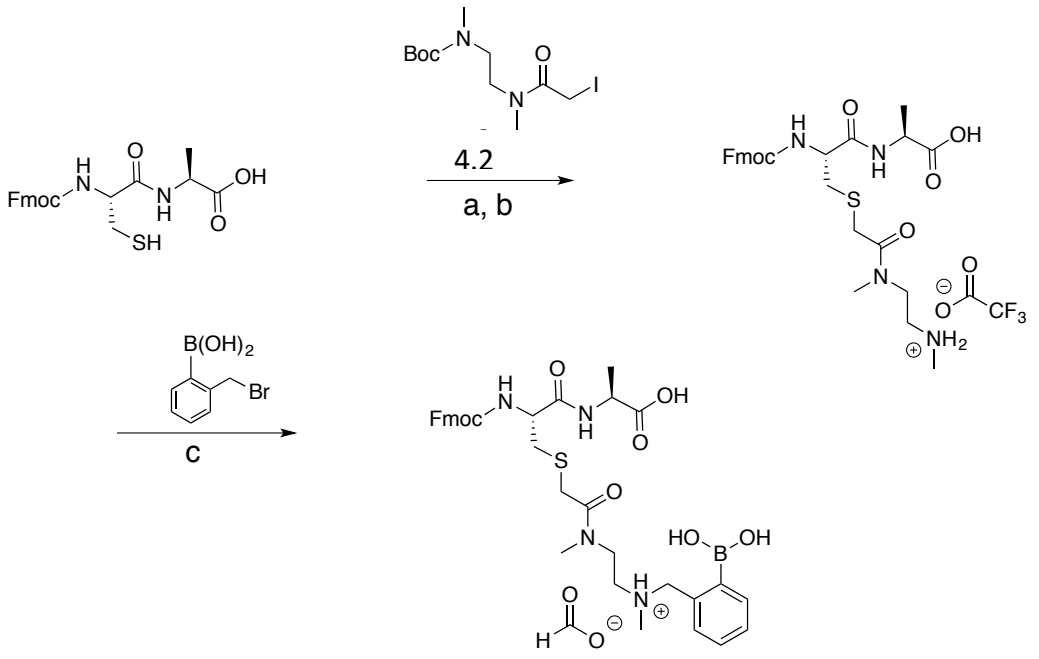
4.3

Scheme 4.1: Synthesis of compound 1.

After successful functionalization of DMEDA as a proof of concept experiment, we set out to devise a linker approach to place secondary amines on peptides. Cysteine residues are reported to be selectively modified with iodoacetamides, therefore a linker with this functional handle provided a method for selectively introducing a secondary amine.[35] The synthesis of this linker involved a single Boc protection of DMEDA (Scheme 4.2). Literature accounts regularly describe mono-protection with diamine compounds by using an excess of the amino compound relative to boc-anhydride.[36] A mixture of products was observed: unreacted starting material, singly protected product, and doubly protected product. Aqueous washes removed unreacted starting material. No further purification was performed because the doubly protected product was inert in the following step. Mono-boc DMEDA was modified with chloroacetyl chloride, and the desired product was isolated. A Finklestein reaction afforded iodoacetamide linker 4.2. Once made, cysteine in Fmoc-CysAla-OH was modified with the linker (Scheme 4.3). Deprotection with TFA solution occurred quantitatively, followed by alkylation with *o*-(bromomethyl)phenylboronic acid (BMPBA)(4.3). Linker 2 should prove widely useful to selectively modify peptides with cysteine residues. [31]



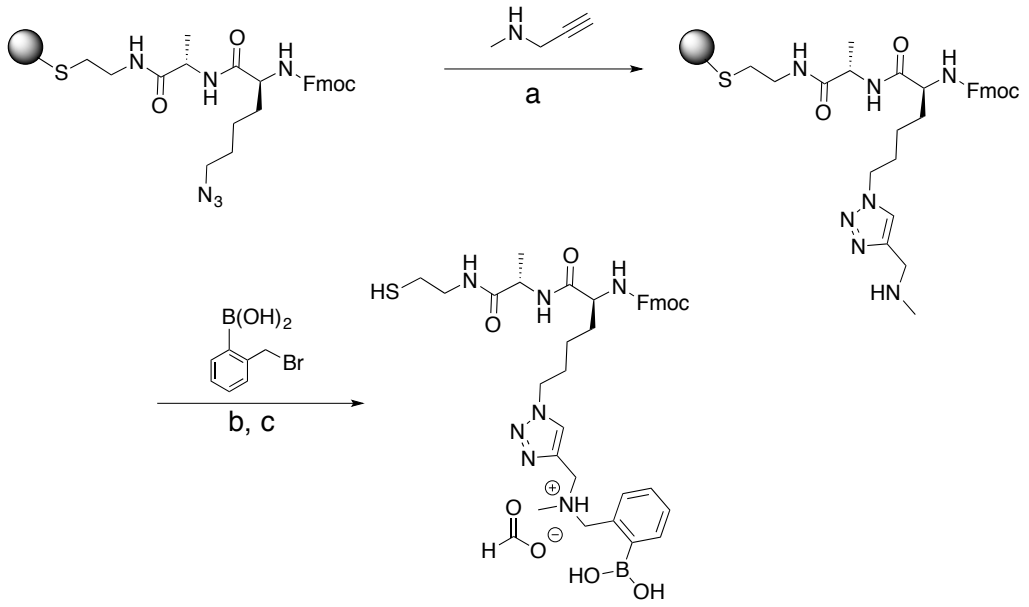
Scheme 4.2: Synthesis of mono-Boc diamine.



Scheme 4.3: a) MeCN, MeOH, TEA, Pyr, H_2O, RT, 4 hrs. b) TFA, TIS, H_2O, RT, 4hrs. c) MeCN, H_2O, RT, overnight.

4.2.2 Solid-phase Incorporation of Secondary Amines and Alkylation

To expand the utility of the alkylation approach, we sought to develop solid-phase reactions for the incorporation of secondary amines. A cysteamine resin and an azido lysine building block also allowed incorporation of a secondary amine (Scheme 4.4). The unnatural azido residue can be modified via click chemistry while immobilized on the solid support.[37] We used N-methyl propargylamine (NMPA) to click on a secondary amine. To ensure the Fmoc protecting group did not get deprotected by the basic amine, repetitions of the click reaction were performed. Three one-hour repetitions were the maximum number before Fmoc deprotection of product became significant.



4.2

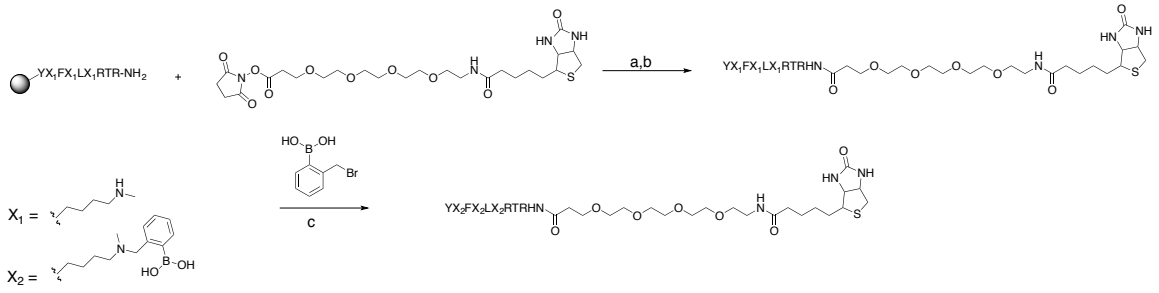
Scheme 4.4: Synthesis of compound 4. a) DMF, CuI, TBTA, Sodium Ascorbate, RT, 1 hr. (3 repetitions). b) MeCN, H_2O, overnight. c) TFA, TIS, H_2O, RT, 4 hrs.

Interestingly, Wang resin did not yield the desired product. However, "clicking" did occur with a cysteamine resin. Upon cleavage, the product (4.4) contained a thiol functional group as the c-terminus, which could be used for solution-phase derivatization if desired.

4.2.3 Building-block Incorporation of Secondary Amines and Alkylation

Using commercially available and synthetically accessible N_ε-methyl lysine as a building block proved to be the most straightforward method for introducing secondary amines. This building block ensured the placement of multiple secondary amines along a sequence with high reliability. Fmoc-Lys(N_ε, Me)A-OH gave the highest purified yield (60%; Scheme 4.5). Only a single modification step post-solid phase synthesis was required to make a boronic acid peptide. Thus, it minimized the cumulative inefficiencies

at each reaction step in the former two approaches. Using this approach, seven complex biotinylated multi-boronic acid peptides were synthesized (Scheme 4.5, Figure 4.2).



Scheme 4.5: Synthesis of compound 6. a) DMF, DIPEA, RT, 2 hrs (2 repetitions) b) TFA, TIS, H_2O, RT, 4hrs. c) MeCN, H_2O, RT, overnight.

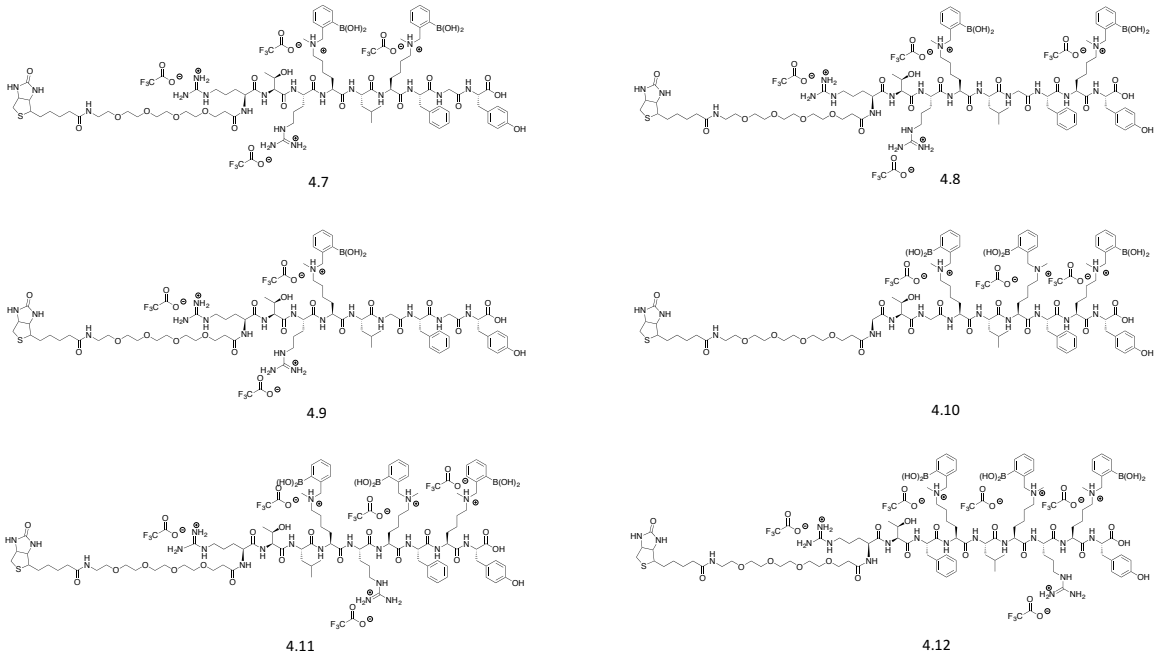


Figure 4.2: Biotinylated boronic acid peptides.

Of the seven biotinylated peptides, **9** had the highest purified yield (45%). The purified yields of peptides with two or three boronic acids ranged from 8%-22%. We attributed the lower yield to the more difficult purification with increasing number of boronic acids on a peptide due to streaking in the purification steps.

The HPLC analytical trace for **9** gave a relatively sharp peak as the product eluted. However, when two or three boronic side chains were present, the elution peak was broadened, suggesting again that the phenyl boronic acid moieties interact with the column. Higher percentages of organic solvent were needed to elute the peptide. The broadening of the elution peak led to low signal intensity during HPLC purification. This broadening led to a low signal intensity during HPLC purification, which made identification of the peaks challenging, and hence poor detection of sample elution was hypothesized to contribute to the lower yields for peptides with more than one boronic acid.

4.3 Conclusions

We developed three different approaches for derivatizing peptides with secondary amines, followed by modification with *o*-(bromomethyl)phenylboronic acid. Unlike most previously published methods for incorporation of boronic acids into peptides, this alkylation strategy allows for easy incorporation of an *o*-aminomethyl functionality on the aryl boronic acid, which has been previously shown to increase binding affinity of boronic acids to saccharides. The first approach involved the synthesis of a linker with an iodoacetamide at one side and boc protected secondary amine on the other end. Cysteine residues were selectively targeted in a model peptide. After deprotection, the secondary amine was alkylated with a boronic acid. In the second approach, N-methyl propargyl amine was used in click chemistry to introduce secondary amines followed by alkylation with a boronic acid in the solid phase. Post cleavage, a boronic acid peptide with a C-terminal thiol group was isolated. Lastly, the commercially available N_{ε}-methyl lysine

building block was used as the most direct method for making boronic acid peptides. It gave the highest yield and was the method of choice to make a set of seven biotinylated peptides. These peptides had up to three boronic acids, although purification became difficult and the associated yields decreased with increasing numbers of boronic acids.

Compound	Sequence	Yield
4.6	$RTRX_2LX_2FX_2Y$	7.5%
4.7	$RTRX_2LX_2FGY$	11%
4.8	$RTRX_2LGFX_2Y$	19%
4.9	$RTRX_2LGFGY$	45%
4.10	$GTGX_2LX_2FX_2Y$	25%
4.11	$RTLX_2RX_2FX_2Y$	17%
4.12	$RTFX_2LX_2RX_2Y$	22%

Table 4.1: Biotinylated peptide library sequences and yields. PEG4-Biotin was placed on the N-terminus. X_2 = $-CH_2CH_2CH_2CH_2(NH(CH_3)(CH_2C_6H_6B(OH)_2))$

Compounds such as 4.6 – 4.12 are currently being investigated as pull-down reagents for cell-surface saccharides. Due to the utility of the synthetic methodologies we report herein we anticipate that others can adopt them for site-specific incorporation of boronic acids in peptides and proteins.

4.4 Experimental

4.4.1 General Materials

For automated Fmoc amino solid-phase peptide synthesis, Ala, Pbf (Arg), Trt (Cys), Gly, Leu, Boc (Lys), Phe, and Boc (Thr) were purchased from P3 biosystems. Fmoc-Lys(N_ε, Me)-OH was purchased from Chem. Pep. Inc. Fmoc-Lys(N_3)-OH was purchased from Chem-Impex, Inc. Cysteamine 2-ClTrt (0.95 mmol/g) resin and Fmoc-Tyr(tBu)-Wang resin (0.46 mmol/g) were purchased from AnaSpec, Inc. Fmoc-Ala-Wang (100-200 mesh, 0.72 mmol/g) was purchased NovaBiochem. DMF, DCM, piperidine used for automated solid-phase peptide synthesis were purchased from Fisher Scientific and Sigma-

Aldrich. N, N'-dimethylethylenediamine, chloroacetyl chloride, and sodium iodide, N-methyl propargyl amine, copper iodide, and sodium ascorbate were purchased from Sigma-Aldrich. *o*-(bromomethyl) phenylboronic acid was purchased from Combi-Blocks. EZ-Link™ NHS-PEG_4-Biotin was purchased from Thermo Scientific.

4.4.2 General Instrumentation

A Prelude peptide synthesizer (Protein Technologies, Inc.) was used for automated-solid phase synthesis of the peptides. For the longer peptides, a Liberty Blue microwave peptide synthesizer was used. Preparative HPLC purification of peptides was performed using an Agilent Zorbax SB-C_{18} Prep HT column 21.2 x 250 mm. Analytical HPLC characterization of peptides was performed using an Agilent Zorbax column 4.6 x 250 mm; 1 ml/min., 5-95% MeCN (0.1 % TFA) in 35 min. (RT). A Gemini C_{18} 3.5 micron 2.1 x 50 mm was used for online separation; 0.7 ml/min., 5-95% MeCN (0.1 % formic acid) in 12 min. (RT). An Agilent Technologies 6530 Accurate Mass QTofLC/MS was used for high-resolution mass spectra of purified peptides. Solvents used were HPLC grade. For small molecule organic compounds, reverse-phase combi-flash was used for purification if necessary.

4.4.3 General Procedure (A): synthesis of Dimeric Peptides and N_ε-methyl Lysine Peptides

Fmoc-CysAla-Wang resin, Fmoc-Lys(N_3)Ala-S-Resin, Fmoc- Lys(N_ε, Me)Ala-Wang, and biotinylated N_ε-methyl lysine peptides resin (100 μmol) were synthesized by automated sequential coupling of N_α-Fmoc-amino acid (0.1 M) in DMF in the presence of *N,N,N,N*-Tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU, 0.15 M) and Hunig's base (0.2 M) with gentle nitrogen bubbling for 30 mins. at room temperature. A total of three repetitions were performed for each amino acid building block incorporated, followed by DMF (3 ml, 3 min., 3x) and DCM (3 ml, 3 min., 3x) washes.

For peptides longer than three amino acid residues, a 0.8 M LiCl wash (3 ml, 3 min., 3x) was included after swelling with DCM. After the synthesis, resin was washed with glacial AcOH (5 mL, 3x), DCM (5 mL, 3x), and MeOH (5 mL, 3x). Resins were placed under vacuum overnight. Fmoc-Cys Ala was cleaved from the resin using trifluoroacetic acid (TFA), triisopropylsilane, 1,2-ethanedithiol (EDT), and nanopure water (94: 1.0: 2.5: 2.5) (4 hrs.) for Fmoc-CysAla-OH. TFA was evaporated and the remaining oil was precipitated with diethyl ether at 0 °C. No further purification of the crude peptide was performed.

4.4.4 General Procedure (B): solution-phase alkylation with *o*-(bromomethyl)phenylboronic acid.

Peptides were dissolved in a mixture of H_2O and MeCN (0.6 mL, 1:1 v/v). To this solution, 0.1 mL of Hunig's base was added. If the peptide precipitated, 0.15 mL of MeOH was added. The solution was further diluted with 0.2 mL of MeCN, followed by addition of 3.5 equivalents of *o*-(bromomethyl)phenylboronic acid. The reaction was allowed to stir overnight at RT. The following day, additional boronic acid (3.5 equivalents) was added, followed by addition of 0.05 mL of Hunig's base. The reaction was incubated at room temperature for an additional 2 hrs. Preparative HPLC was used to purify the peptides. Purified samples were placed on the rotary evaporator to remove MeOH. The aqueous remnants were frozen and lyophilized overnight.

4.4.5 General Procedure (C): solid-phase modification of the N-terminus with EZ-Link™ NHS-PEG_4-Biotin: Post automated solid phase synthesis.

Fmoc group was removed with 3 mL of piperidine (20% v. in DMF, 3 min., 3x) followed by washes with DMF (3 mL, 3 min., 3x), DCM (3 mL, 3 min., 3x), and 0.8 M LiCl (3 mL, 3 min., 3x). A 1.5 mL solution of EZ-Link™ NHS-PEG_4-Biotin was introduced

with 0.5 mL solution of 1.2 M Hunig's base. Gentle stirring under nitrogen was performed for 1 hr. at RT, followed by washes with DMF (3 ml, 3 min, 3x), DCM (3 ml, 3 min., 3x), and 0.8 M LiCl (3 ml, 3 min., 3x). Coupling of biotin to peptide was repeated a total of three times. Resins were washed with glacial AcOH (5 mL, 3x), DCM (5 mL, 3x), and MeOH (5 mL, 3x), and resins were placed under vacuum overnight. Peptides were cleaved with trifluoroacetic acid (TFA), triisopropylsilane, and nanopure water (95: 2.5: 2.5) (4 hrs.), followed by removal of TFA, and precipitated with diethyl ether at 0°C. No further purification of the crude peptide was performed.

4.4.6 Synthesis of (((ethane-1,2-diylbis(methylazanediyl))bis(methylene))bis(2,1-phenylene))diboronic acid (4.1*)*.

N, N'-dimethylethylenediamine (2.0 mmol) was dissolved in a solution of MeCN, MeOH, DMF (1.4 mL, 71/14/14 (v/v/v)), followed by addition of 2-bromomethyl)phenylboronic (4.0 mmol). The reaction was stirred overnight at RT. The compound was purified via reverse-phase combi-flash. Purified yield: 16%. HRMS-ESI$^+$ (MeOH/H_2O): calcd.: m/z = 319.20310; found: m/z = 319.20060 $[M\text{-}2H_2O+H]^+$. ^{1}H NMR (400 MHz, Acetonitrile-d_3) δ 7.72 (dt, J = 7.4, 1.0 Hz, 1H), 7.66 – 7.63 (m, 1H), 7.61 – 7.58 (m, 1H), 7.38 – 7.23 (m, 6H), 5.00 (d, J = 0.8 Hz, 1H), 4.86 (s, 2H), 4.52 (s, 1H), 3.89 (s, 7H), 3.30 (d, J = 0.8 Hz, 2H), 2.78 – 2.70 (m, 1H).

Synthesis of tert-butyl (2-(2-iodo-N-methylacetamido)ethyl)(methyl)carbamate (4.2)— Boc-anhydride (0.011 mol) was dissolved in 20 mL of DCM and added dropwise to a solution of N, N' dimethylethylenediamine (0.034 mol) in 30 mL of DCM under N_2. The reaction was stirred overnight at RT. The DCM was removed via rotary evaporation. The residual oil was washed with EtOAc, with water, and brine. Ethyl acetate was evaporated and a colorless oil remained. No further purification was performed. The crude product (1g) was dissolved

in 20 mL of dry DCM and placed in an ice bath, and chloroacetyl chloride (0.0058 mmol) was introduced slowly under N_2. The solution turned dark brown. The reaction was allowed to come to RT and stirred overnight. DCM was removed via rotary evaporation. The remaining oil was dissolved in EtOAc and washed with distilled water and brine (3x). The organic layer was dried with Na_2SO_4. After filtration, EtOAc was removed by rotary evaporation. The compound was purified by normal-phase combi-flash (SI). Isolated product (0.11 mol) was dissolved in 1 mL of acetone followed by the introduction of NaI (0.22 mol). Precipitate formed after ten minutes. The reaction was stirred overnight at RT. The acetone was removed via rotary evaporation, followed by three washes with brine. No further purification was required. Yield 73%. LRMS-ESI/APC$^{+/-}$ (MeOH/H_2O): calcd.: *m/z* = 379.2; found *m/z* = 379.05 $[M+Na]^+$.

Synthesis of Fmoc-C(B)A (4.3)—Fmoc-CA-OH (0.07 mmol) was dissolved in 0.3 mL of H_2O, followed by addition of 0.1 mL of MeOH:TEA:Pyr:H_2O (7:1 :1:1) (v/v/v). A solution of compound **3** (0.07 mmol) in 0.5 mL MeCN:H_2O (69:31) (v/v) was introduced. The boc protecting group was removed using trifluoroacetic acid (TFA), triisopropylsilane, and nanopure water (95: 2.5: 2.5) (4 hrs.). TFA was removed, and the product was precipitated with diethyl ether at 0°C. No further purification of the peptide was performed. The peptide was alkylated with *o*-(bromomethyl)phenylboronic acid following general procedure B. The peptide was purified via preparative HPLC. MeOH was removed by rotary evaporation, followed by freezing and lyophilization of aqueous remnants. Purified yield: 6%. HRMS-ESI$^+$ (MeOH/H_2O): calcd.: *m/z* = 658.27420; found: *m/z* = 658.27390 $[M-2H_2O+H]^+$. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.30 (s, 1H), 7.82 (d, *J* = 7.6 Hz, 2H), 7.65 (t, *J* = 6.8 Hz, 2H), 7.38 (t, *J* = 7.4 Hz, 2H), 7.33 – 7.25 (m, 3H), 7.16 (d, *J* = 7.3 Hz, 1H), 4.25 (t, *J* = 6.6 Hz, 2H), 4.20 – 4.11 (m, 1H), 3.94 (d, *J* = 7.0 Hz, 1H), 3.88 (s, 4H), 3.72 (d, *J* = 10.9 Hz, 1H), 3.57 – 3.39 (m, 2H), 3.40 – 3.30 (m, 1H), 2.91 (dd,

J = 13.8, 5.2 Hz, 1H), 2.71 (d, *J* = 14.0 Hz, 3H), 2.68 – 2.58 (m, 4H), 2.24 (s, 2H), 2.16 (s, 1H), 1.18 (dd, *J* = 7.0, 1.7 Hz, 3H). ^{13}C NMR (126 MHz, DMSO-d6) δ 170.31, 156.90, 144.37, 144.36, 144.25, 141.36, 135.41, 135.40, 130.34, 130.14, 128.60, 128.02, 126.00, 120.83, 66.76, 63.25, 63.23, 54.97, 54.97, 52.75, 50.19, 47.26, 47.26, 44.12, 41.44, 40.45, 40.45, 40.06, 40.02, 39.90, 39.87, 39.84, 39.73, 39.70, 39.67, 39.56, 39.53, 39.51, 39.40, 39.37, 39.34, 39.23, 39.20, 39.17, 39.04, 39.01, 35.87, 34.48, 34.12, 18.92, 18.91.

Synthesis of Fmoc-Lys(N_3,B)Ala-SH (4.4). Fmoc-Lys(N_3)Ala-S-Resin. Mixing of solvents, preparation of click solutions, and click reaction was done under N_2. Resin (100 μmol) and alkyne (1 eq) were combined with 1 mL of DMF followed by the addition of 0.2 mL solution of CuI (0.5 eq), tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (**TBTA**) (1 eq), and sodium ascorbate (1 eq) in DMF and stirred for 1 hr. at RT. The reaction solution was drained from the resin and fresh reaction solution was introduced two more times. The resin was washed with DMF (3 mL, 3 min., 3x), DCM (3 mL, 3 min., 3x), glacial AcOH (5 mL, 3x), DCM (5 mL, 3x), and MeOH (5 mL, 3x). The resin was then placed under vacuum overnight, followed by cleavage using trifluoroacetic acid (TFA), triisopropylsilane, 1,2-ethanedithiol (EDT), and nanopure water (94: 1.0: 2.5: 2.5) (4 hrs.). TFA was evaporated and remaining oil was precipitated with diethyl ether at 0°C. Peptide was purified via preparative HPLC. MeOH was removed by rotary evaporation, followed by freezing and lyophilization of aqueous remnants. Purified yield: 14%. HRMS-ESI$^+$ (MeOH/H_2O): calcd.: *m/z* = 749.32520; found: *m/z* = 749.32680 $[M+Na]^+$. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.30 (s, 1H), 7.97 (d, *J* = 6.7 Hz, 1H), 7.84 (t, *J* = 6.8 Hz, 2H), 7.65 (q, *J* = 6.2 Hz, 2H), 7.58 (d, *J* = 7.1 Hz, 1H), 7.39 (q, *J* = 7.0 Hz, 2H), 7.29 (q, *J* = 7.5, 7.0 Hz, 2H), 7.22 (dd, *J* = 8.7, 6.6 Hz, 1H), 7.15 (q, *J* = 6.8 Hz, 1H), 4.30 (t, *J* = 6.8 Hz, 2H), 4.22 (dd, *J* = 7.4, 4.0 Hz, 2H), 4.16 (td, *J* = 6.9, 3.3 Hz, 2H), 3.93 (dd, *J* = 9.4, 4.9 Hz, 1H), 3.65 (d, *J* = 4.4 Hz, 2H), 3.47 (d, *J* = 5.2 Hz, 0H), 3.33 (dt, *J* = 13.5, 6.8 Hz, 1H), 3.25 (dt, *J* =

13.4, 6.6 Hz, 1H), 2.78 – 2.63 (m, 2H), 2.03 (d, *J* = 4.5 Hz, 3H), 1.77 (d, *J* = 9.3 Hz, 3H), 1.64 (s, 1H), 1.52 (d, *J* = 11.8 Hz, 1H), 1.35 – 1.19 (m, 4H), 1.16 (d, *J* = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, DMSO-d_6) δ 172.78, 172.08, 156.53, 144.13, 144.03, 143.25, 141.86, 141.07, 134.48, 129.44, 129.21, 129.09, 128.16, 127.57, 127.19, 125.65, 124.80, 121.78, 120.51, 72.44, 66.14, 61.71, 60.55, 54.72, 49.83, 49.70, 48.63, 47.02, 42.26, 40.28, 40.11, 40.02, 39.94, 235.31, 175.67, 174.39, 173.56, 173.06, 172.72, 172.43, 171.79, 171.67, 171.22, 167.42, 164.16, 160.27, 160.01, 157.35, 157.30, 155.95, 142.27, 138.02, 135.25, 134.31, 131.05, 130.22, 129.87, 129.81, 129.19, 128.94, 128.54, 127.66, 127.24, 127.23, 118.79, 116.42, 115.59, 110.38, 74.96, 70.39, 70.37, 70.31, 70.18, 69.57, 67.28, 67.03, 62.00, 60.24, 60.13, 58.19, 56.29, 56.15, 54.43, 53.62, 52.39, 49.34, 42.70, 41.00, 40.48, 40.28, 40.17, 39.25, 37.84, 37.10, 36.54, 35.91, 31.41, 29.10, 29.08, 28.92, 28.67, 25.97, 25.43, 24.87, 24.21, 23.63, 23.47, 23.47, 23.34, 23.33, 22.07, 20.17.

Fmoc-Lys(N_ε,Me,B)A-OH (4.5). General procedure B was used for 0.054 mmol of starting material. Purified yield: 60%. Positive HRMS-ESI+ (MeOH/H_2O): calcd: m/z = 610.27010; found: m/z = 610.2685 $[M + Na]^+$. Negative HRMS-ESI (MeOH/H_2O): calcd: m/z = 586.2736; found: m/z = 586.2727 [M H]. ^{1}H NMR (400 MHz, DMSO-d_6) d 8.23 (s, 1H), 7.84 (dd, J = 7.6, 3.1 Hz, 2H), 7.68–7.61 (m, 2H), 7.50 (dt, J = 7.1, 1.1 Hz, 1H), 7.39 (tt, J = 7.5, 1.6 Hz, 2H), 7.35–7.30 (m, 2H), 7.30–7.26 (m, 1H), 7.23 (ddd, J = 7.2, 5.3, 3.4 Hz, 1H), 7.18 (d, J = 7.2 Hz, 1H), 4.63 (s, 2H), 4.28–4.15 (m, 3H), 4.04 (q, J = 7.2 Hz, 1H), 3.93 (dd, J = 9.0, 5.1 Hz, 1H), 3.66 (q, J = 7.3 Hz, 2H), 3.50–3.45 (m, 1H), 3.43–3.37 (m, 1H), 2.19–2.12 (m, 3H), 1.61 (dt, J = 13.8, 7.0 Hz, 1H), 1.49 (h, J = 7.6 Hz, 2H), 1.20 (dd, J = 9.4, 7.1 Hz, 4H).

Biotin-ArgThrArgLys(N_ε,Me,B)LeuLys(N_ε,Me,B)PheLys(N_ε,Me,B)- Tyr-OH (4.6). General procedure B was used for 0.022 mmol of starting material. Purified yield: 7.5%. HRMS-ESI+ (MeOH/H_2O): calcd: 701.0625; found: m/z = 701.0638 [M - 3H_2O +

3H]$^{3+}$. 1 H NMR (500 MHz, DMSO-d_6) d 8.29 (s, 8H), 7.65–7.50 (m, 3H), 7.33–7.03 (m, 18H), 6.93 (d, J = 7.8 Hz, 3H), 6.58 (d, J = 7.9 Hz, 3H), 4.51–4.30 (m, 2H), 3.57 (q, J = 8.5, 7.5 Hz, 5H), 3.46 (d, J = 8.9 Hz, 9H), 3.37 (t, J = 5.7 Hz, 2H), 3.24 (s, 1H), 3.20–3.11 (m, 2H), 3.04 (dt, J = 26.2, 7.3 Hz, 6H), 2.91 (s, 5H), 2.83–2.63 (m, 7H), 2.42–2.24 (m, 10H), 2.23–2.00 (m, 12H), 1.70–1.32 (m, 29H), 1.29–1.05 (m, 14H), 1.04–0.90 (m, 3H), 0.78 (dtd, J = 33.8, 18.2, 17.3, 5.5 Hz, 9H).

Biotin-ArgThrArgLys(N_ε,Me,B)LeuLys(N_ε,Me,B)PheGlyTyr-OH (4.7). General procedure B was used for 0.021 mmol of starting material. Purified yield: 11%. HRMS-ESI+ (MeOH/H_2O): calcd: m/z = 951.0221; found: m/z = 951.0258 [M - 3H_2O + 2H]$^{2+}$. HRMS-ESI (MeOH/H_2O): calcd: m/z = 949.0075; found: m/z = 949.0051 [M - 2H2O 2H]2. 1 H NMR (499 MHz, DMSO-d_6) d 8.35 (s, 8H), 8.16 (s, 4H), 7.59 (s, 3H), 7.20–7.05 (m, 4H), 6.94 (q, J = 8.8 Hz, 2H), 6.54 (d, J = 8.3 Hz, 1H), 4.34–4.27 (m, 1H), 4.13 (dd, J = 7.9, 4.3 Hz, 1H), 3.42–3.35 (m, 3H), 2.91 (s, 1H), 2.81–2.65 (m, 1H), 2.57 (d, J = 12.6 Hz, 2H), 2.45–2.17 (m, 1H), 2.12–1.99 (m, 13H), 1.69 (d, J = 17.0 Hz, 1H), 1.63–1.55 (m, 1H), 1.54–1.34 (m, 18H), 1.33–1.22 (m, 1H), 1.18 (d, J = 7.1 Hz, 5H), 1.00 (td, J = 14.0, 12.8, 6.7 Hz, 4H), 0.86 (d, J = 6.1 Hz, 1H), 0.80 (t, J = 5.6 Hz, 4H), 0.74 (dd, J = 10.5, 4.6 Hz, 1H). 13C NMR (126 MHz, DMSO-d_6) d 142.07, 133.86, 129.20, 129.17, 127.81, 127.04, 74.53, 57.89.

Biotin-ArgThrArgLys(N_ε,Me,B)LeuGlyPheLys(N_ε,Me,B)Tyr-OH (4.8). General procedure B was used for 0.015 mmol of starting material. Purified yield: 19%. HRMS-ESI$^+$ (MeOH/H_2O): calcd: m/z = 634.6855; found: m/z = 634.6832 [M - 2H_2O + 3H]$^{3+}$. ^{1}H NMR (500 MHz, DMSO-d_6) d 8.34 (s, 9H), 7.84 (t, J = 5.5 Hz, 1H), 7.66 (s, 1H), 7.37–6.87 (m, 3H), 6.58 (d, J = 8.4 Hz, 3H), 6.41 (s, 1H), 6.36 (s, 1H), 4.32–4.28 (m, 1H), 4.12 (d, J = 7.2 Hz, 1H), 3.99 (d, J = 24.1 Hz, 1H), 3.21–3.14 (m, 2H), 3.10–2.89 (m, 1H), 2.85–2.78 (m, 2H), 2.39–2.29 (m, 1H), 2.21 (q, J = 0.5 Hz, 2H), 1.63–1.35 (m, 6H), 1.33–

1.09 (m, 3H), 1.01 (dd, J = 15.0, 6.0 Hz, 4H), 0.84 (ddd, J = 23.6, 11.2, 5.7 Hz, 8H). 13C NMR (126 MHz, DMSO-d6) d 179.31, 172.13, 165.36, 162.71, 157.24, 129.20, 127.96, 69.75, 63.36, 61.04, 59.20, 55.40, 39.12, 35.09, 30.67, 30.65, 28.17, 28.02, 25.24.

Biotin-ArgThrArgLys(N_ε,Me,B)LeuGlyPheGlyTyr-OH (4.9). General procedure B was used for 0.015 mmol of starting material. Purified yield: 45%. HRMS-ESI$^+$ (MeOH/H_2O): calcd: m/z = 567.6386; found: m/z = found 567.6403 $[M - H_2O + 3H]^{3+}$. HRMS-ESI (MeOH/H_2O): calcd: m/z = 848.9370; found: m/z = 848.9370 $[M - H_2O + 2H]^{2+}$. 1 H NMR (500 MHz, DMSO-d6) d 8.35 (s, 8H), 8.13 (s, 1H), 7.92–7.57 (m, 2H), 7.55 (d, J = 7.2 Hz, 1H), 7.38 (d, J = 7.6 Hz, 1H), 7.34 (d, J = 5.8 Hz, 0H), 7.32–7.26 (m, 3H), 7.26–7.18 (m, 7H), 7.18–7.02 (m, 2H), 6.94 (d, J = 8.3 Hz, 3H), 6.59 (d, J = 7.9 Hz, 3H), 6.39 (d, J = 28.5 Hz, 2H), 4.88 (s, 1H), 4.76 (s, 1H), 4.64 (d, J = 4.2 Hz, 1H), 4.58 (s, 2H), 4.44 (s, 0H), 4.36 (s, 1H), 4.32–4.27 (m, 1H), 4.20 (s, 3H), 4.13 (s, 2H), 4.02 (s, 2H), 3.88–3.70 (m, 3H), 3.59 (d, J = 6.4 Hz, 2H), 3.49 (d, J = 2.6 Hz, 10H), 3.39 (t, J = 6.0 Hz, 2H), 3.31 (s, 1H), 3.28 (d, J = 5.1 Hz, 5H), 3.22–3.15 (m, 3H), 3.04 (s, 6H), 2.94 (d, J = 12.8 Hz, 1H), 2.82 (dd, J = 12.5, 5.1 Hz, 2H), 2.74 (s, 2H), 2.58 (d, J = 12.4 Hz, 1H), 2.46–2.28 (m, 2H), 2.22–2.20 (m, 2H), 1.47 (d, J = 25.5 Hz, 20H), 1.35–1.16 (m, 2H), 1.02 (d, J = 8.1 Hz, 4H), 0.83 (d, J = 22.7 Hz, 8H). ^{13}C NMR (126 MHz, DMSO-d_6) d 172.45, 172.13, 171.90, 171.66, 171.53, 171.15, 170.33, 170.15, 168.71, 168.71, 167.56, 166.37, 162.73, 157.29, 157.29, 155.45, 143.52, 142.89, 142.27, 138.01, 137.98, 134.00, 133.99, 133.73, 133.25, 130.13, 129.09, 128.93, 128.56, 128.33, 128.05, 127.37, 126.95, 126.81, 126.19, 126.04, 125.72, 125.63, 125.35, 114.66, 73.89, 73.37, 73.33, 73.24, 73.20, 69.76, 69.70, 69.67, 69.55, 69.45, 69.15, 66.76, 66.34, 61.05, 59.22, 58.44, 57.64, 57.59, 57.52, 57.49, 57.47, 55.88, 55.40, 51.87, 40.78, 40.34, 39.45, 39.13,38.45, 37.19, 35.85, 35.09, 30.97, 30.79, 30.65, 30.47, 28.66, 28.18, 28.03, 25.25, 24.73, 24.12, 22.97, 21.60, 21.47, 20.08, 19.76, 18.69.

Biotin-GlyThrGlyLys(N_ε,Me,B)LeuLys(N_ε,Me,B)PheLys(N_ε,Me,B)-Tyr-OH (4.10). General procedure B was used for 0.015 mmol ofstarting material. Purified yield: 25%. HRMS-ESI$^+$ (MeOH/H_2O): calcd: m/z = 635.3418; found: m/z = 635.3437 [M - H_2O + 3H]$^{3+}$.HRMS-ESI (MeOH/H_2O): calcd: m/z = 950.9938; found: m/z =949.9907 [M - $3H_2O$ + 2H]$^{2+}$. ^{1}H NMR (500 MHz, DMSO-d_6) d 8.30 (s, 6H), 7.64 (d, J = 7.2 Hz, 3H), 7.34–7.19 (m, 4H), 7.15 (d,J = 8.8 Hz, 12H), 6.94 (d, J = 7.5 Hz, 3H), 6.59 (d, J = 7.8 Hz, 3H), 4.96 (s, 0H), 4.52 (d, J = 1.8 Hz, 1H), 4.49–4.39 (m, 1H), 4.31 (t, J = 6.4 Hz, 1H), 4.13 (dd, J = 7.8, 4.3 Hz, 2H), 4.00 (t, J = 5.3 Hz, 1H), 3.38 (t, J = 5.9 Hz, 2H), 3.25 (d, J = 1.7 Hz, 1H), 3.20–3.13 (m, 2H), 3.11–2.98 (m, 1H), 2.93 (d, J = 9.7 Hz, 1H), 2.83–2.67 (m, 2H), 2.57 (d, J = 12.5 Hz, 1H), 2.50 (p, J = 1.8 Hz, 11H), 2.45–2.31 (m, 6H), 2.30–2.21 (m, 0H), 2.21–2.17 (m, 1H), 1.94 (s, 1H), 1.66–1.56 (m, 1H), 1.55–1.33 (m, 13H), 1.27 (q, J = 7.8 Hz, 1H), 1.22 (dd, J = 6.6, 2.2 Hz, 9H), 1.12 (d, J = 19.8 Hz, 0H), 1.07–0.94 (m, 4H), 0.93–0.68 (m, 6H). 13C NMR (126 MHz, DMSO-d_6) d 234.84, 234.82, 180.21, 172.94, 171.36, 170.97, 169.88, 169.35, 165.77, 163.23, 155.70, 142.16, 141.70, 135.07, 133.88, 130.93, 130.55, 129.64, 129.64, 129.41, 129.07, 129.07, 128.30, 127.61, 127.16, 126.84, 115.00, 109.87, 74.38, 72.40, 70.05, 69.99, 69.93,69.84, 69.81, 69.32, 69.32, 66.89, 66.72, 63.79, 62.93, 61.35,60.46, 59.50, 57.78, 55.79, 55.72, 53.45, 42.42, 41.85, 40.32, 40.14,38.71, 36.16, 35.40, 31.77, 31.29, 31.11, 30.97, 30.95, 30.91, 30.87, 30.84, 30.80, 30.77, 30.71, 30.61, 30.46, 28.51, 28.31, 25.55, 25.15, 24.43, 23.26, 21.79, 19.82, 17.80, 12.84, 4.89, 15.08, 15.08.

Biotin-ArgThrLeuLys(N_ε,Me,B)ArgLys(N_ε,Me,B)PheLys(N_ε,Me,B)-Tyr-OH (4.11). General procedure B was used for 0.022 mmol of starting material. Purified yield: 17%. HRMS-ESI$^+$ (MeOH/H_2O): calcd: m/z = 701.3950; found: m/z = 701.3926 [M - H_2O + 3H]$^{3+}$. HRMS-ESI (MeOH/H_2O): calcd: m/z = 1050.0735; found: m/z = 1050.0684 [M - 3H2O + 2H]$^{2+}$. ^{1}H NMR (500 MHz, DMSO-d_6) d 8.28 (s, 5H), 7.64 (dt, J = 19.1, 8.7 Hz,

3H), 7.41–6.99 (m, 18H), 6.94 (d, J = 8.0 Hz, 2H), 6.60 (d, J = 8.0 Hz, 2H), 4.59–4.41 (m, 2H), 4.39–4.28 (m, 1H), 4.11 (d, J = 14.8 Hz, 97H), 3.56 (q, J = 14.4, 10.9 Hz, 2H), 3.45 (d, J = 7.9 Hz, 12H), 3.38 (d, J = 6.0 Hz, 2H), 3.16 (t, J = 6.3 Hz, 2H), 3.10–2.96 (m, 6H), 2.84–2.61 (m, 7H), 2.57 (d, J = 12.7 Hz, 1H), 2.45–2.10 (m, 10H), 2.05 (t, J = 7.4 Hz, 2H), 1.79–1.30

Synthesis of Biotin-ArgThrPheLys(N_ε, Me, B)LeuLys(N_ε, Me, B)ArgLys(N_ε, Me, B)Tyr-OH (4.12). General procedure B was used for 0.022 mmol of starting material. Purified yield: 22 %. Positive HRMS-ESI$^+$ (MeOH/H_2O): calcd.: *m/z* = 700.73010; found: *m/z* = found 700.73090 $[M - H_2O + 3H]^{3+}$. HRMS-ESI$^-$ (MeOH/H_2O): calcd.: *m/z* = calcd. 1048.57700; found: *m/z* = 1048.56120 $[M-3H_2O-2H]^{2-}$. ^{1}H NMR (500 MHz, DMSO-d_6) δ 8.28 (s, 8H), 7.65 – 7.57 (m, 3H), 7.31 – 7.20 (m, 11H), 7.15 (t, *J* = 11.0 Hz, 16H), 6.90 (d, *J* = 8.1 Hz, 0H), 6.85 (d, *J* = 8.1 Hz, 2H), 6.62 – 6.50 (m, 3H), 6.37 (d, *J* = 19.1 Hz, 0H), 4.51 (s, 1H), 4.49 (s, 1H), 4.31 (dd, *J* = 7.8, 5.0 Hz, 2H), 4.29 – 4.22 (m, 0H), 4.22 – 4.09 (m, 2H), 3.98 (s, 2H), 3.65 – 3.51 (m, 4H), 3.50 – 3.46 (m, 12H), 3.44 (s, 4H), 3.37 (t, *J* = 5.8 Hz, 3H), 3.24 (d, *J* = 1.3 Hz, 1H), 3.16 (t, *J* = 5.9 Hz, 3H), 3.11 – 2.87 (m, 4H), 2.79 (dd, *J* = 12.6, 5.2 Hz, 3H), 2.20 – 2.17 (m, 1H), 1.95 – 1.91 (m, 1H), 1.82 – 1.73 (m, 0H), 1.59 (s, 7H), 1.54 – 1.32 (m, 26H), 1.28 (p, *J* = 7.1 Hz, 1H), 1.18 (d, *J* = 18.2 Hz, 8H), 0.95 (d, *J* = 6.3 Hz, 4H), 0.81 (dd, *J* = 22.7, 6.6 Hz, 9H).

4.5 Acknowledgements

We thank our following funding sources: the National Institute of Health (5DP1GM106408), the Defense Advanced Research Projects Agency (DARPA, N66001-14-2-4051), and the Welch Regents Chair (F-0046).

We also thank Anslyn undergraduate reserachers Saumil Datar, Cedric Ginestra, and Kristi Morris for resynthesizing compounds for NMR characterization.

4.6 Experimental Characterization Data

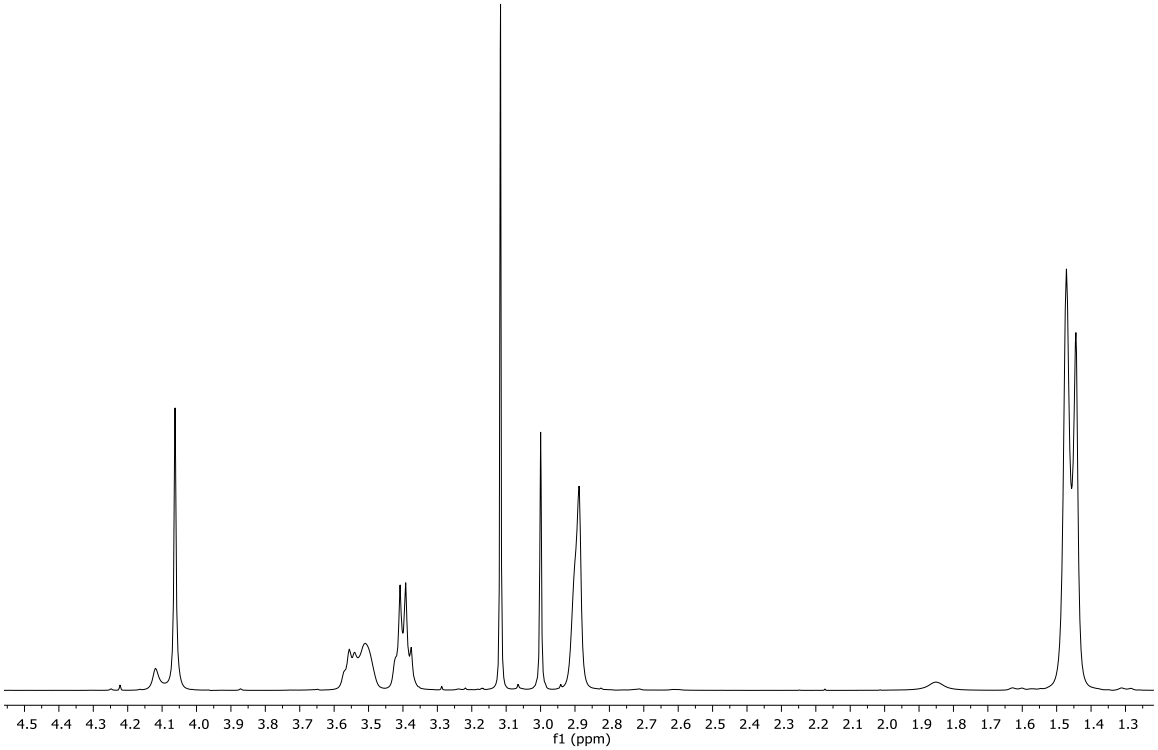
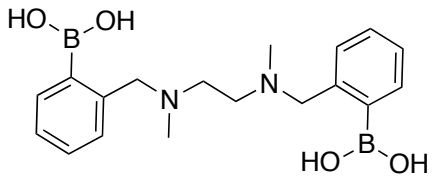


Figure 4.3. Proton NMR spectra for compound *tert*-butyl (2-(2-chloro-*N*-methylacetamido)ethyl)(methyl)carbamate.



1

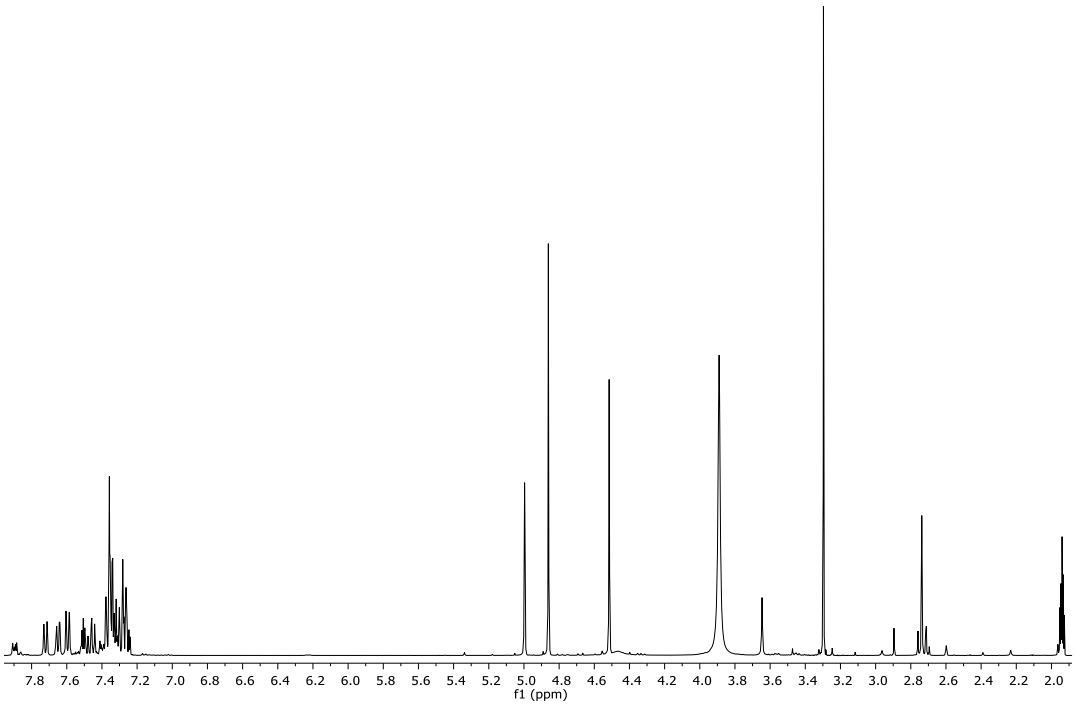
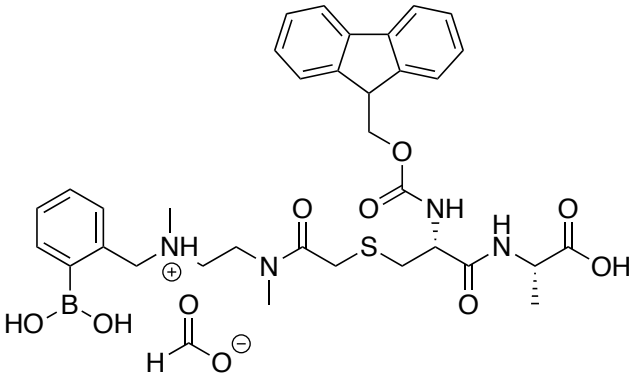


Figure 4.4: Proton NMR spectra for compound 4.1.



3

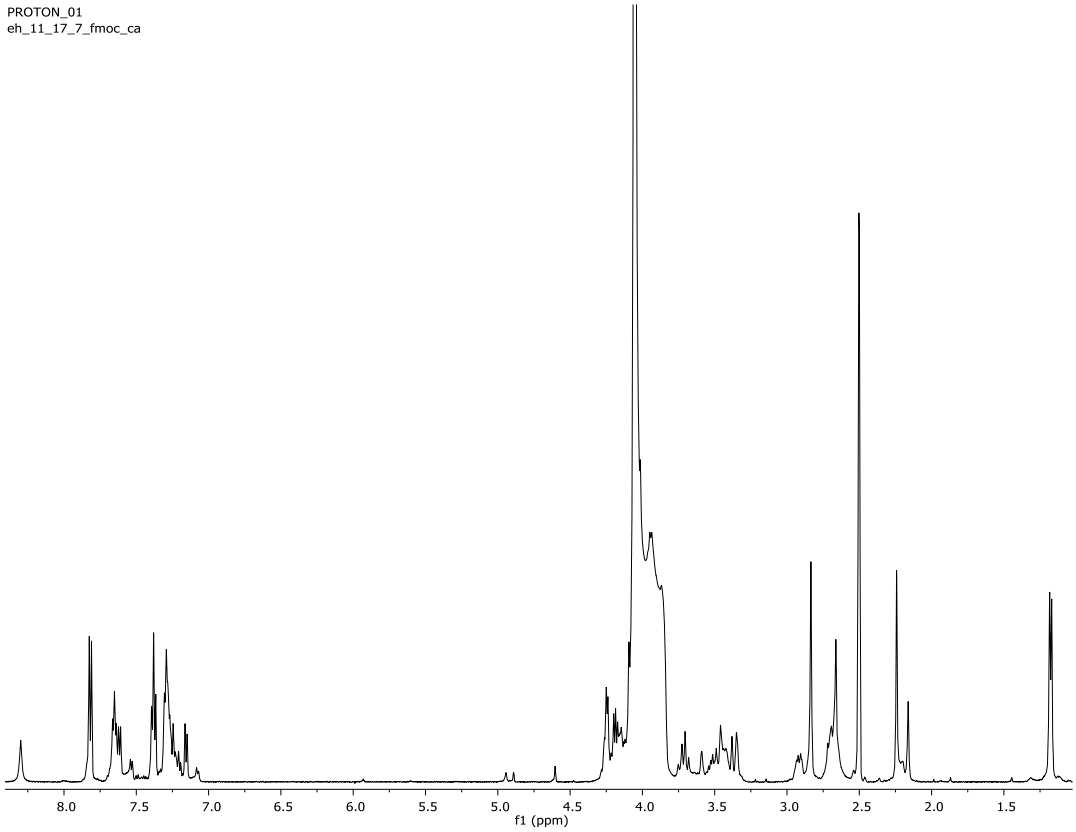


Figure 4.5: Proton NMR spectra for compound 4.3.

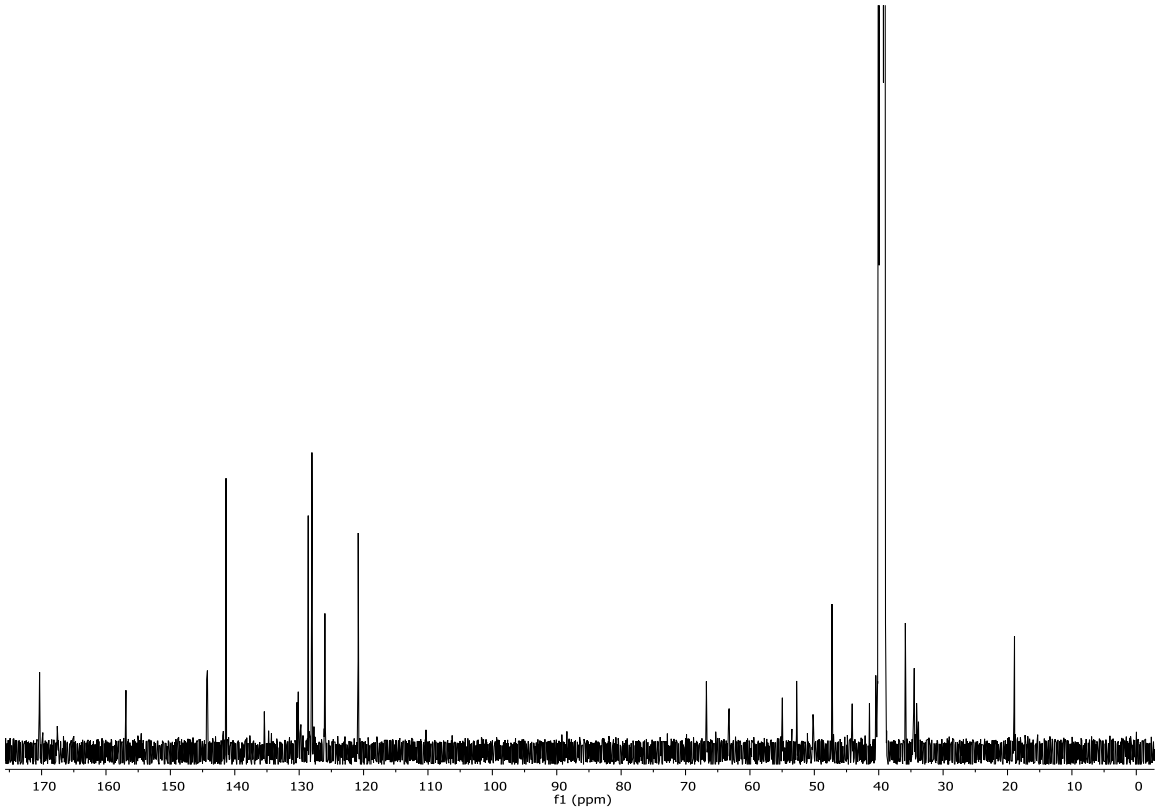
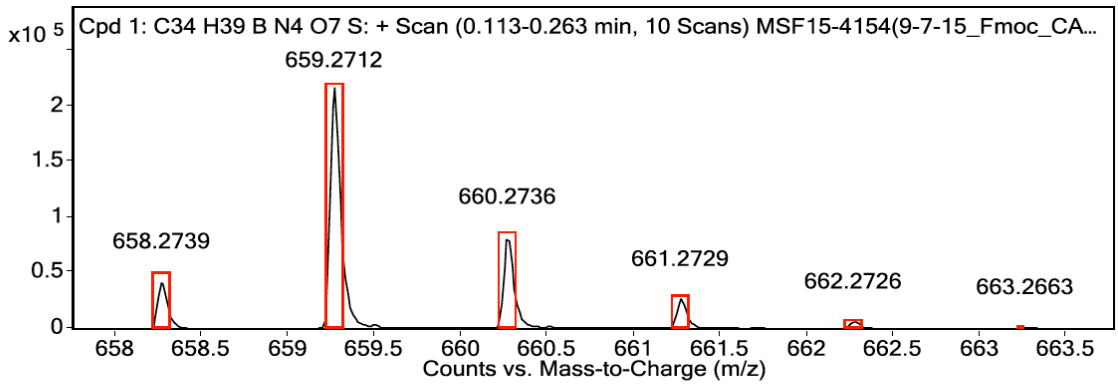
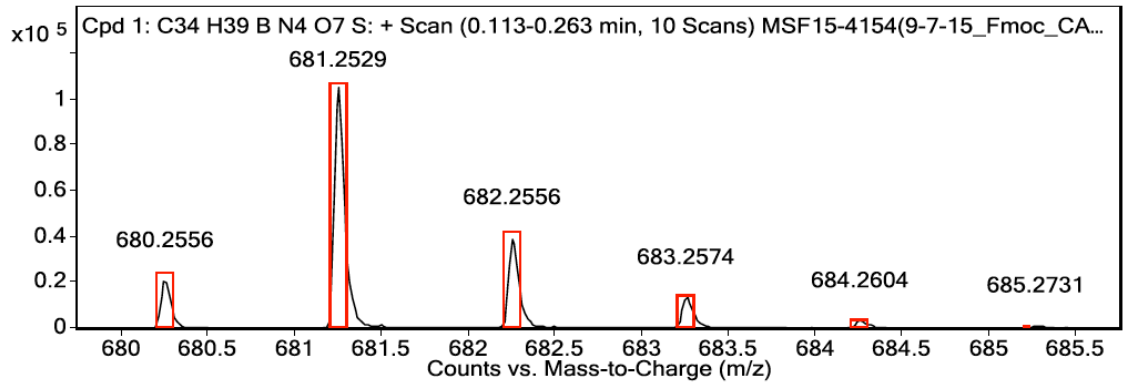


Figure 4.6: Carbon NMR spectra for compound 4.3.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
365.17110			1587199.29			
658.27390	658.27420	1	42899.02	C34H39BN4O7S	(M+H)+	0.34
659.27120	659.27110	1	218568.75	C34H39BN4O7S	(M+H)+	-0.17
660.27360	660.27390	1	83138.85	C34H39BN4O7S	(M+H)+	0.43
661.27290	661.27330	1	27086.4	C34H39BN4O7S	(M+H)+	0.68
662.27260	662.27420	1	6774.03	C34H39BN4O7S	(M+H)+	2.42
663.26630	663.27550	1	1685.74	C34H39BN4O7S	(M+H)+	13.87
664.27650	664.27720	1	1210.82	C34H39BN4O7S	(M+H)+	1.09
665.28400	665.27910	1	1771.52	C34H39BN4O7S	(M+H)+	-7.41



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
365.17110			1587199.29			
680.25560	680.25610	1	21376.07	C34H39BN4O7S	(M+Na)+	0.79
681.25290	681.25310	1	106548.02	C34H39BN4O7S	(M+Na)+	0.21
682.25560	682.25580	1	40203.08	C34H39BN4O7S	(M+Na)+	0.28
683.25740	683.25530	1	13898.59	C34H39BN4O7S	(M+Na)+	-3.1
684.26040	684.25610	1	3726.04	C34H39BN4O7S	(M+Na)+	-6.25
685.27310	685.25750	1	1567.39	C34H39BN4O7S	(M+Na)+	-22.79
686.29420	686.25920	1	578.84	C34H39BN4O7S	(M+Na)+	-51.11

Figure 4.7: HRMS data for compound 4.3.

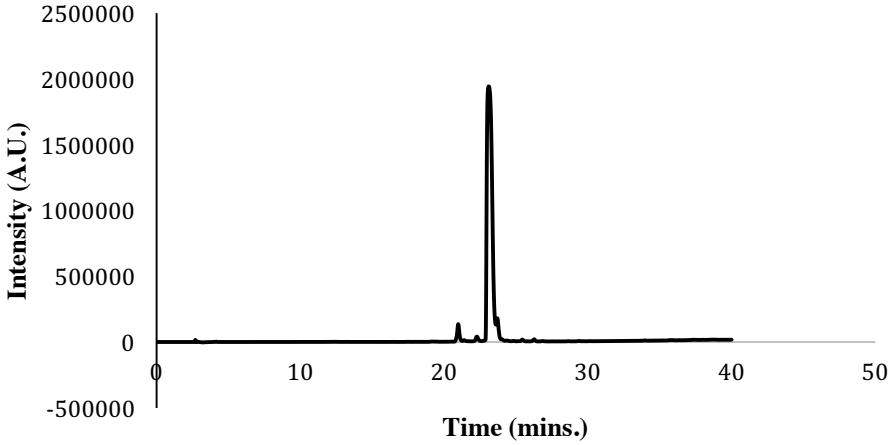
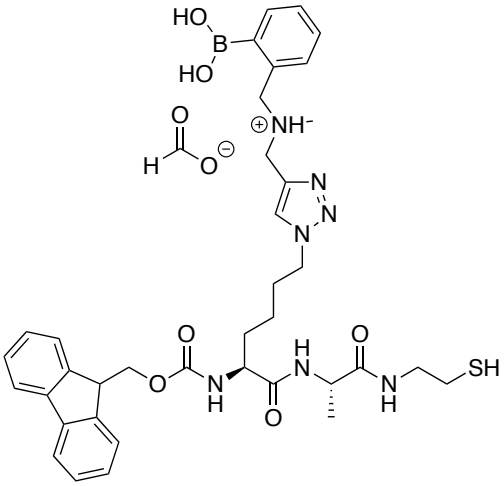


Figure 4.8: Purity check for compound 4.3. Retention time: 21.952 min.



4.4

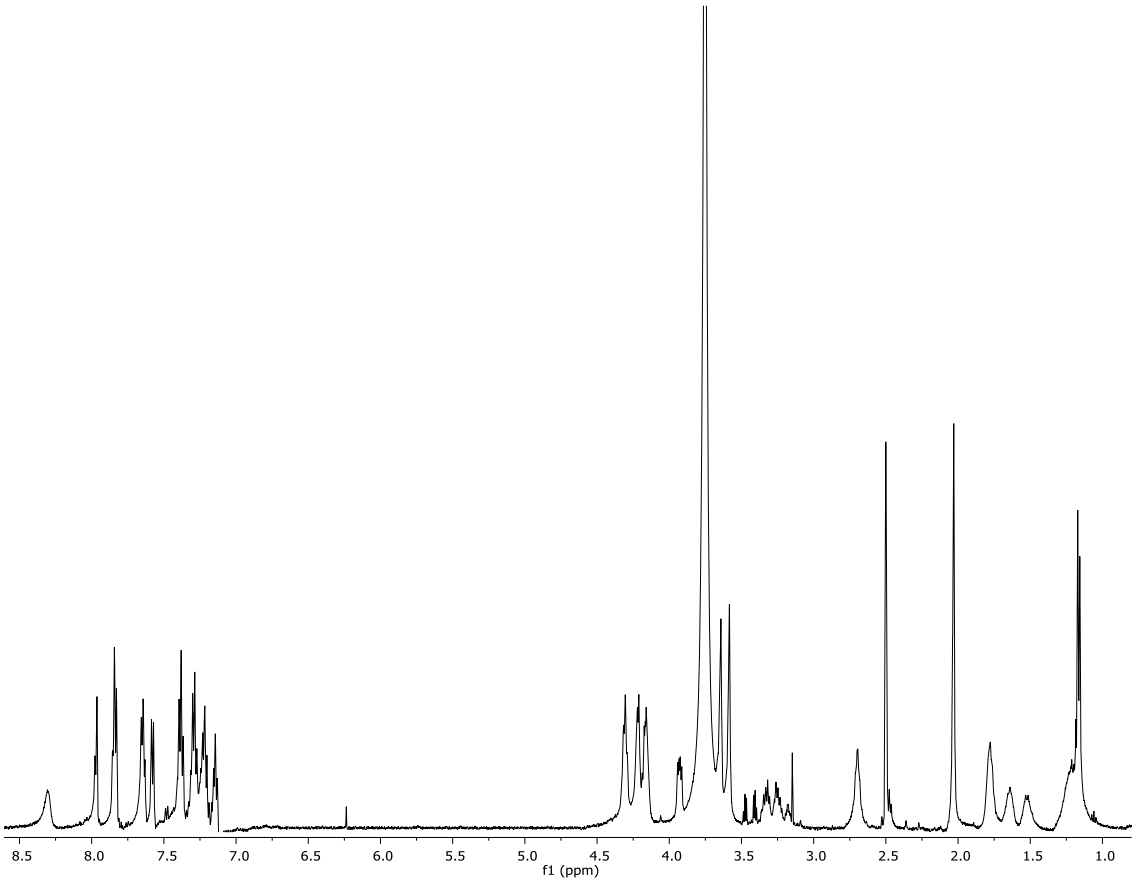


Figure 4.9: Proton NMR spectra for compound 4.4.

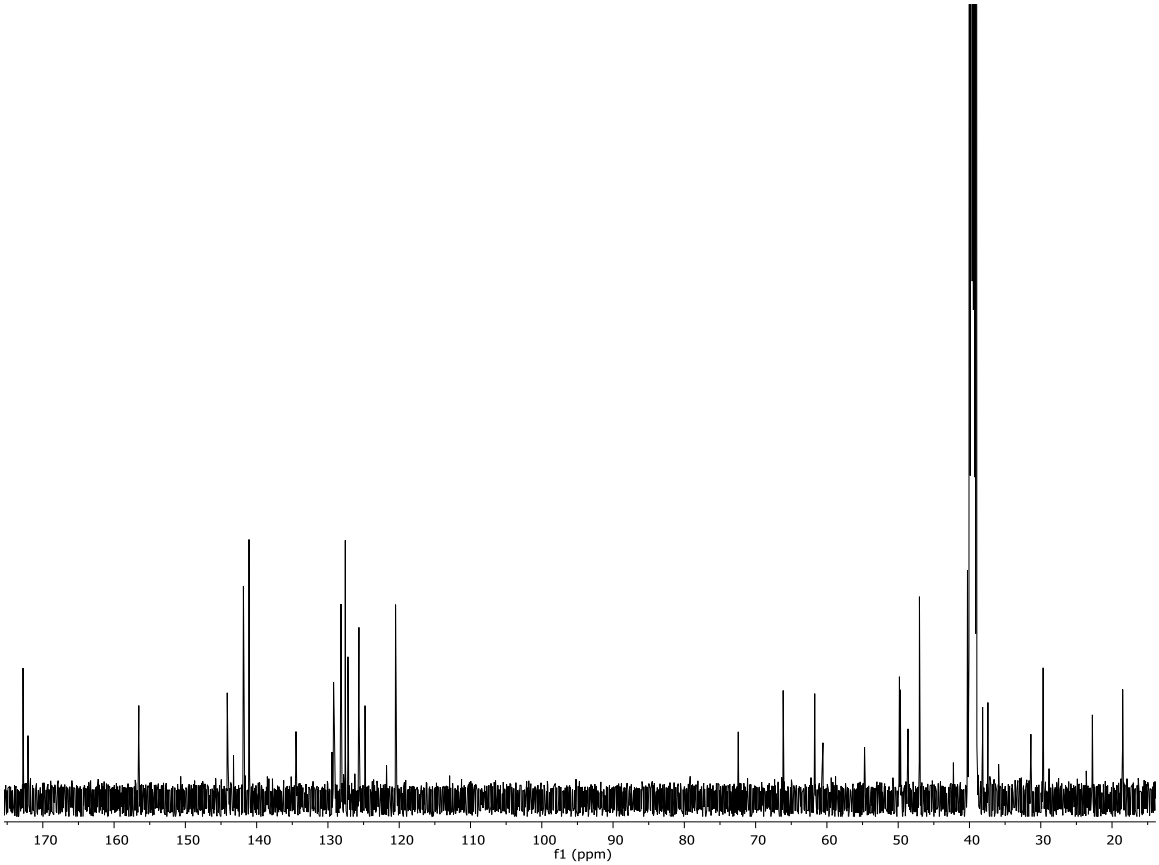
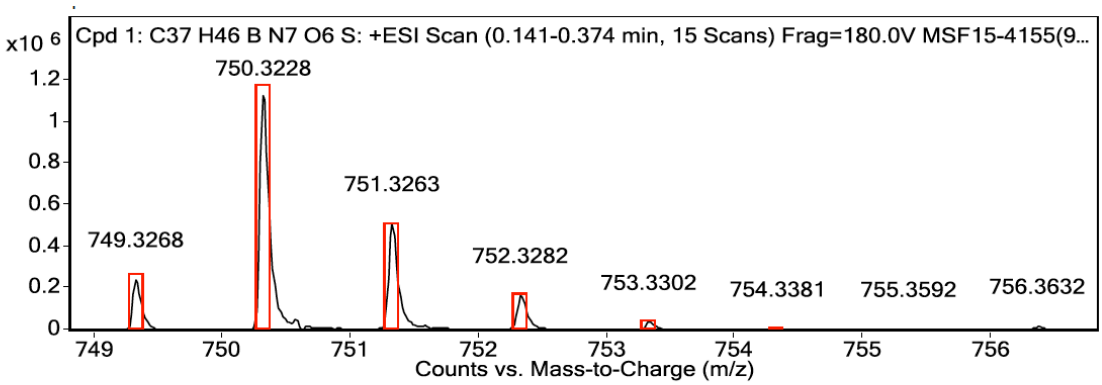


Figure 4.10: Carbon NMR spectra for compound 4.4.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
710.32960			3891242.44			
749.32680	749.32520	1	245938.14	C37H46BN7O6S	(M+Na)+	-2.22
750.32280	750.32220	1	1161121.49	C37H46BN7O6S	(M+Na)+	-0.77
751.32630	751.32480	1	514350.97	C37H46BN7O6S	(M+Na)+	-2
752.32820	752.32460	1	169860.33	C37H46BN7O6S	(M+Na)+	-4.73
753.33020	753.32530	1	40307.77	C37H46BN7O6S	(M+Na)+	-6.57
754.33810	754.32660	1	7863.18	C37H46BN7O6S	(M+Na)+	-15.22
755.35920	755.32820	1	6798.21	C37H46BN7O6S	(M+Na)+	-41.06
756.36320	756.33000	1	21715.75	C37H46BN7O6S	(M+Na)+	-43.87

Figure 4.11: HRMS for compound 4.4.

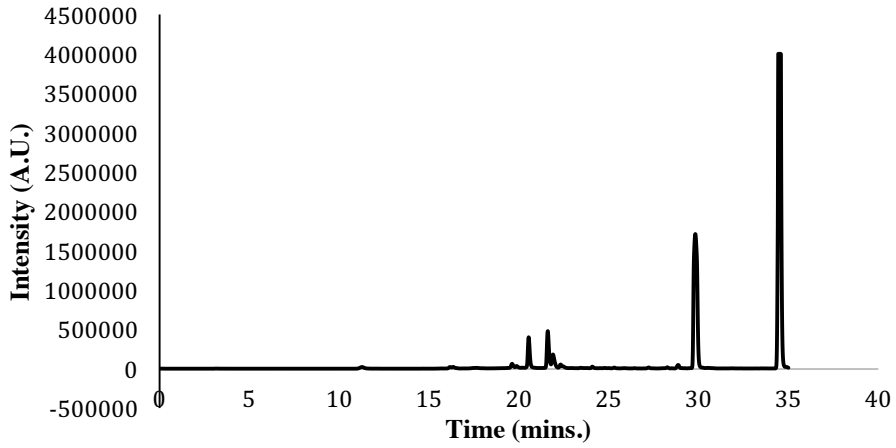
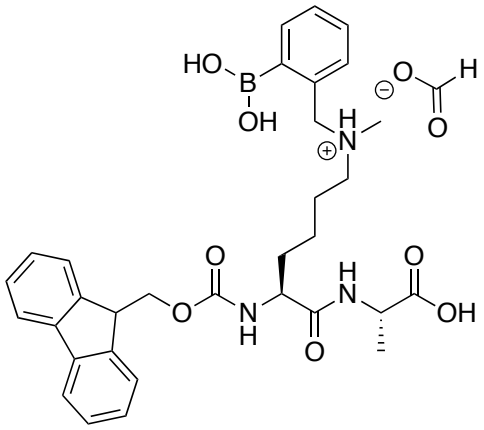


Figure 4.12: Purity check for compound 4.4. Retention time: 34.443 min.



4.5

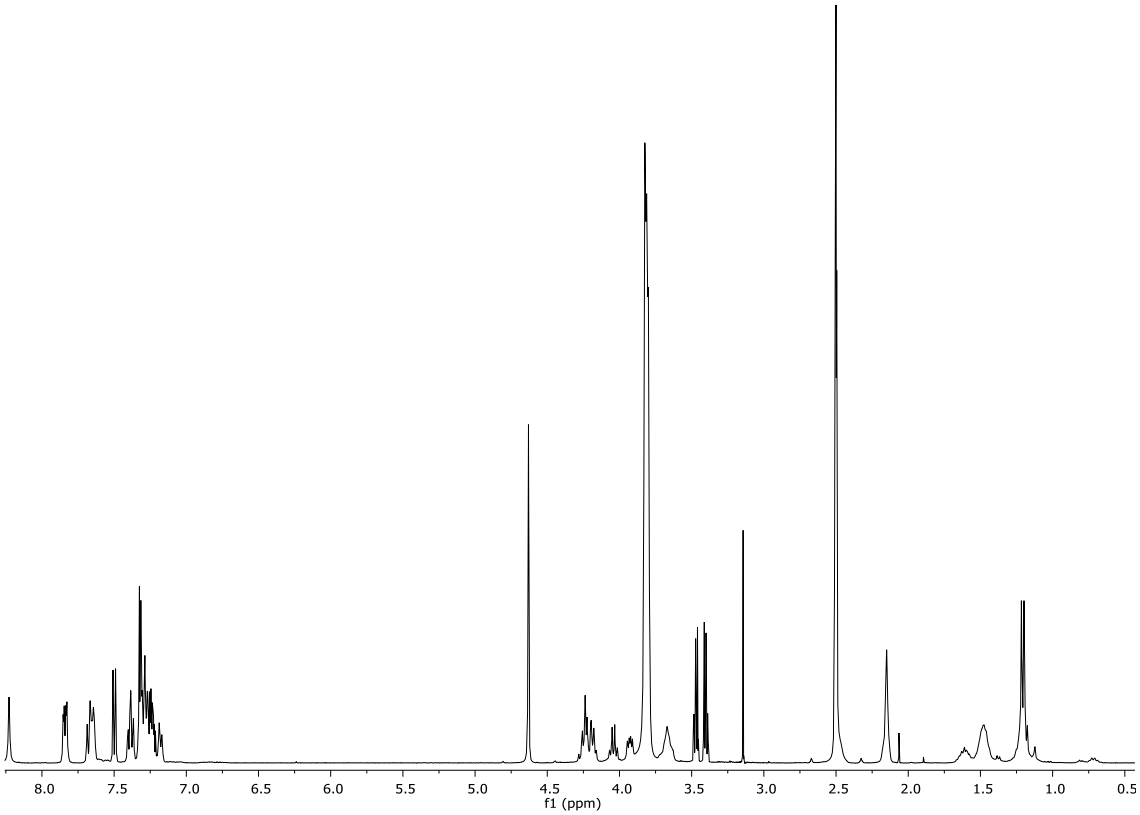
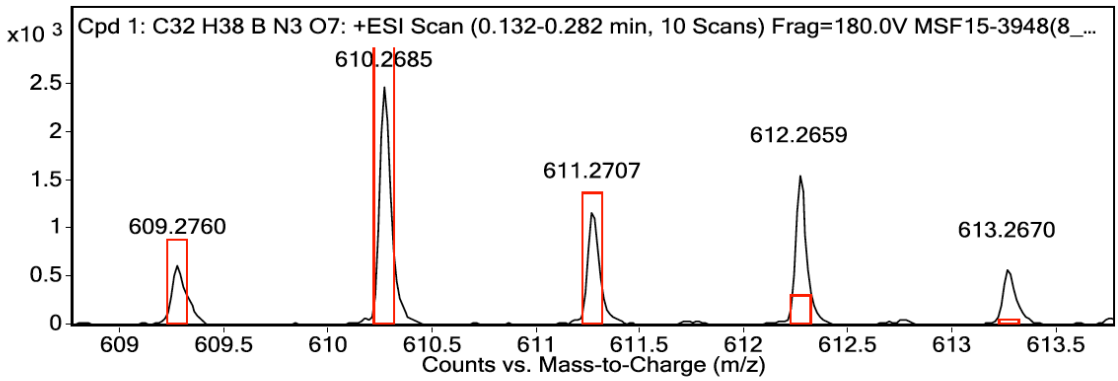
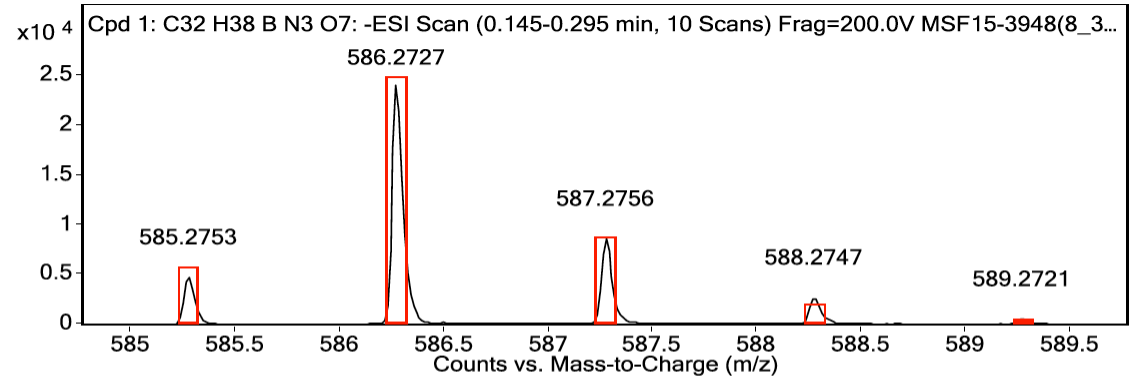


Figure 4.13: Proton NMR spectra for compound 4.5.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
570.27720			29566.74			
609.27600	609.27310	1	614.45	C32H38BN3O7	(M+Na)+	-4.71
610.26850	610.27010	1	2476.66	C32H38BN3O7	(M+Na)+	2.62
611.27070	611.27300	1	1195.69	C32H38BN3O7	(M+Na)+	3.78
612.26590	612.27580	1	1566.68	C32H38BN3O7	(M+Na)+	16.11
613.26700	613.27840	1	579.2	C32H38BN3O7	(M+Na)+	18.65



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
585.27530	585.27660	1	4800.18	C32H38BN3O7	(M-H)-	2.35
586.27270	586.27360	1	24297.9	C32H38BN3O7	(M-H)-	1.48
587.27560	587.27650	1	8683.08	C32H38BN3O7	(M-H)-	1.67
588.27470	588.27930	1	2756.03	C32H38BN3O7	(M-H)-	7.8
589.27210	589.28190	1	588.39	C32H38BN3O7	(M-H)-	16.68
590.27610	590.28450	1	182.32	C32H38BN3O7	(M-H)-	14.21

Figure 4.14: HRMS data for compound 4.5.

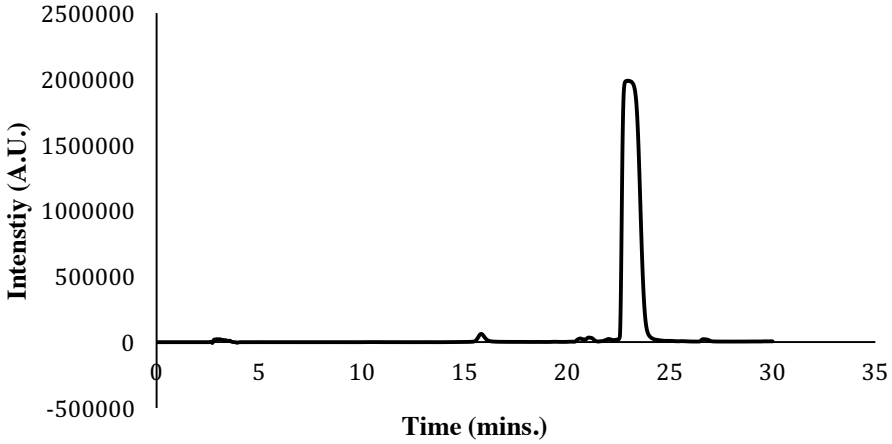
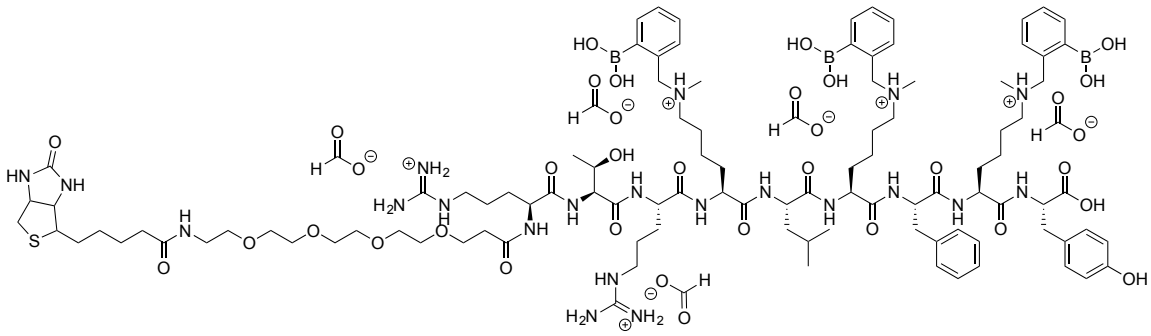


Figure 4.15: Purity check for compound 4.5. Retention time: 22.944 min.



4.6

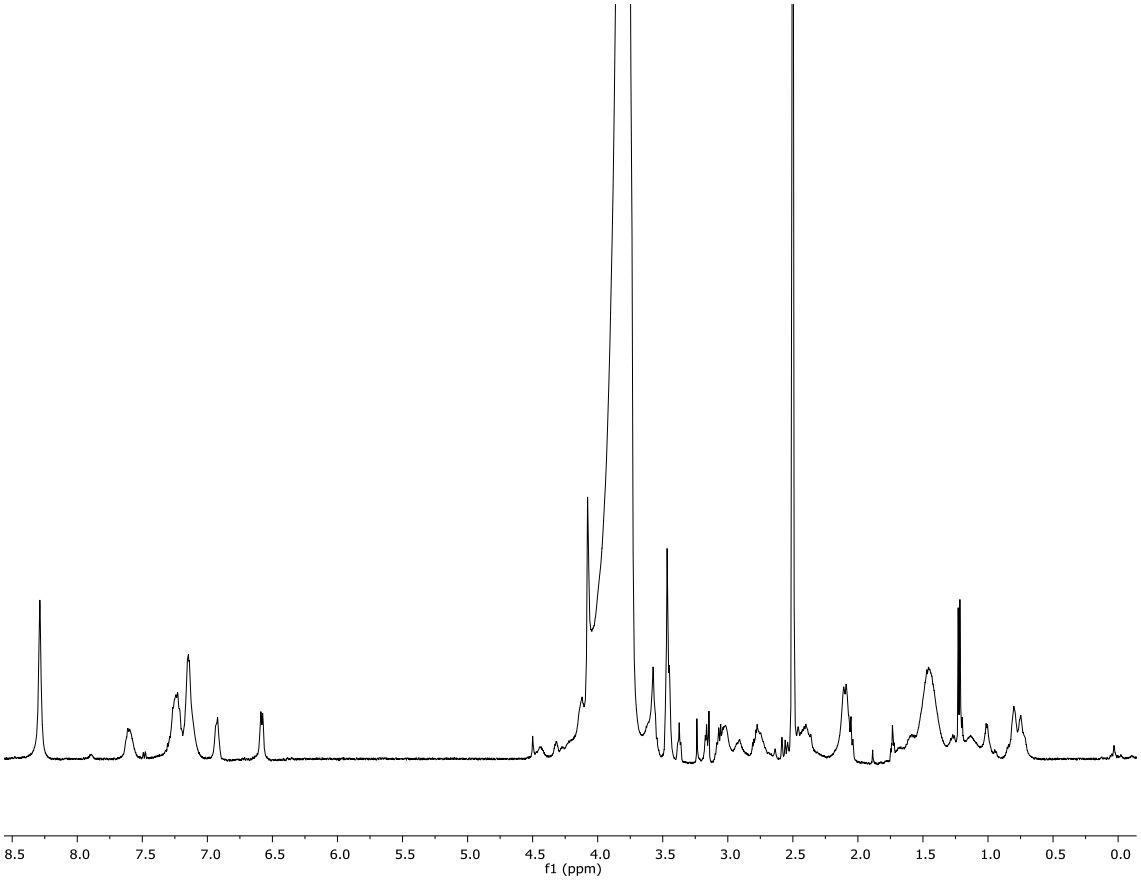
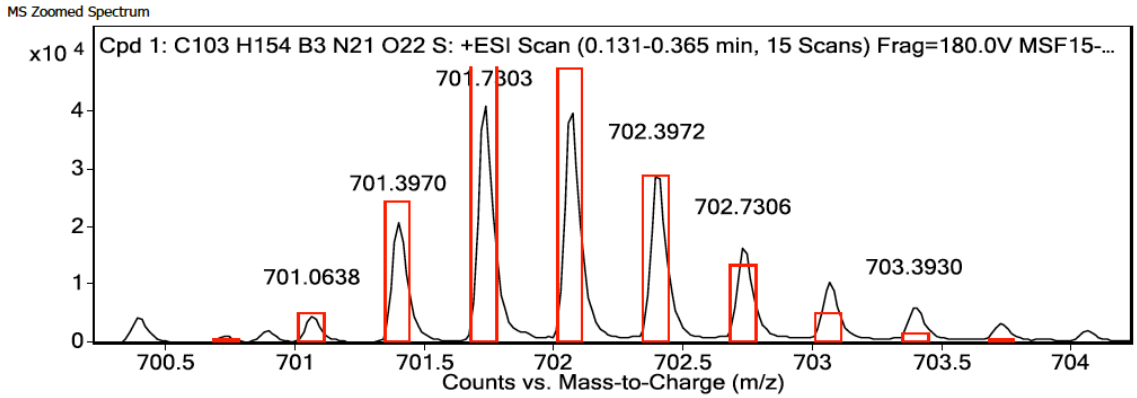


Figure 4.16: Proton NMR spectra for compound 4.6.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
189.06950			123801.57			
700.72720	700.73010	3	1229.77	C103H154B3N21O22S	(M+3H)+3	4.25
701.06380	701.06250	3	4590.75	C103H154B3N21O22S	(M+3H)+3	-1.95
701.39700	701.39500	3	20871.28	C103H154B3N21O22S	(M+3H)+3	-2.91
701.73030	701.72780	3	41272.51	C103H154B3N21O22S	(M+3H)+3	-3.59
702.06390	702.06160	3	40435.42	C103H154B3N21O22S	(M+3H)+3	-3.21
702.39720	702.39550	3	29841.61	C103H154B3N21O22S	(M+3H)+3	-2.4
702.73060	702.72940	3	16789.51	C103H154B3N21O22S	(M+3H)+3	-1.58
703.06180	703.06330	3	10529.62	C103H154B3N21O22S	(M+3H)+3	2.2
703.39300	703.39730	3	6336.79	C103H154B3N21O22S	(M+3H)+3	6.06
703.72440	703.73120	3	3439.32	C103H154B3N21O22S	(M+3H)+3	9.68
704.05510	704.06520	3	2118.36	C103H154B3N21O22S	(M+3H)+3	14.3

--- End Of Report ---

Figure 4.17: HRMS data for compound 4.6.

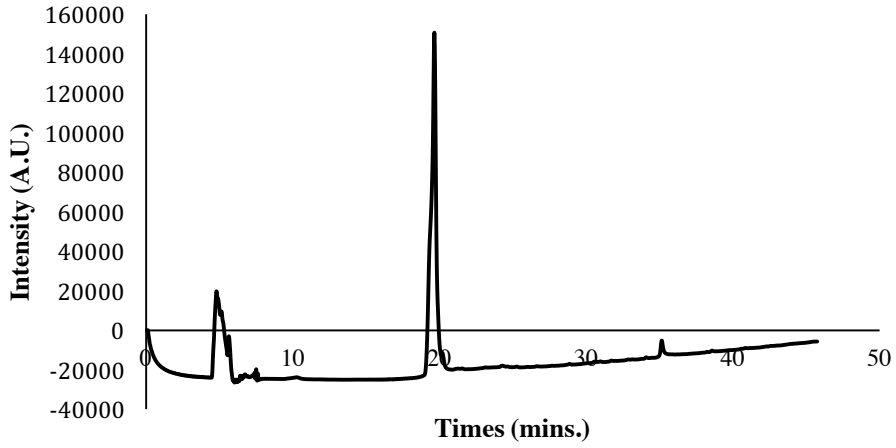
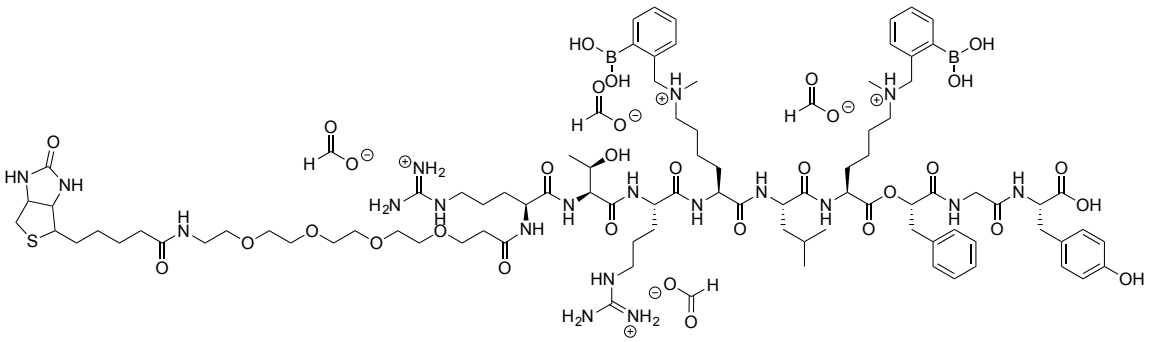


Figure 4.18: Purity check for compound 4.6. Retention time: 19.648 min.



4.7

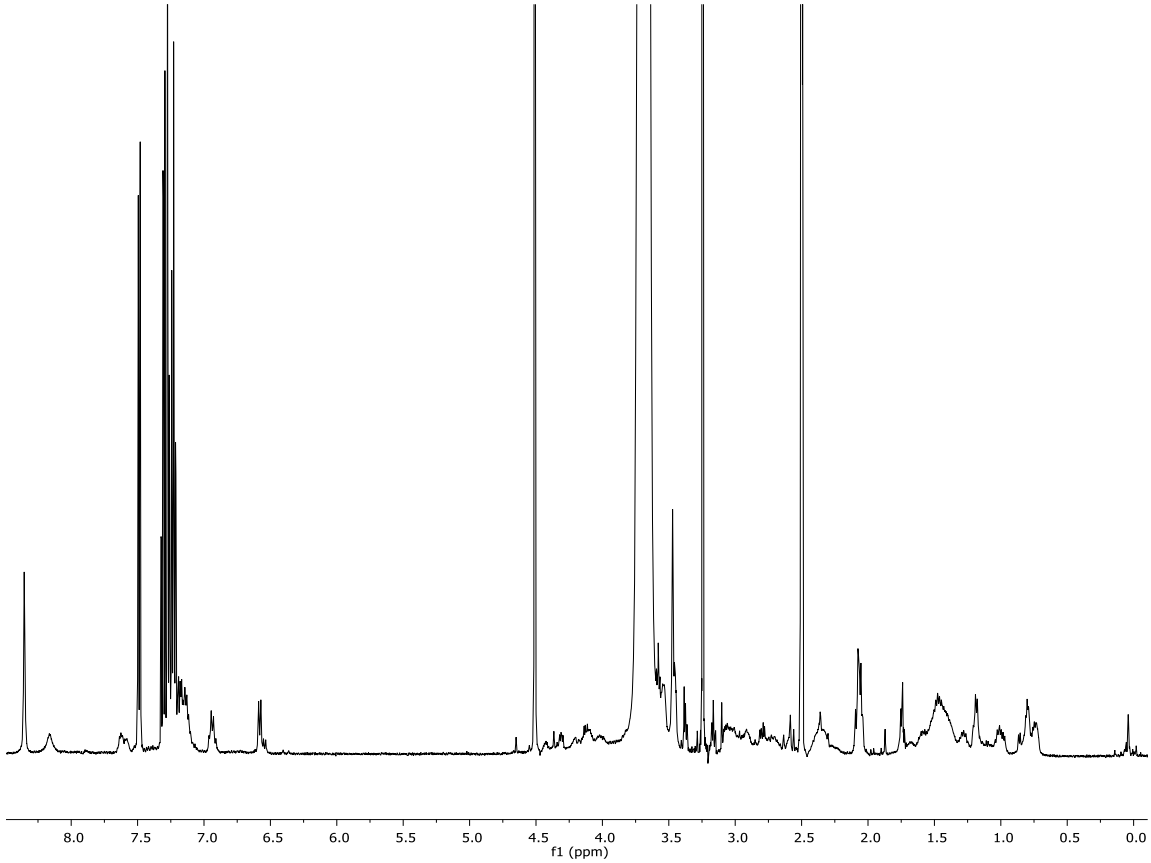


Figure 4.19: Proton NMR spectra for compound 4.7.

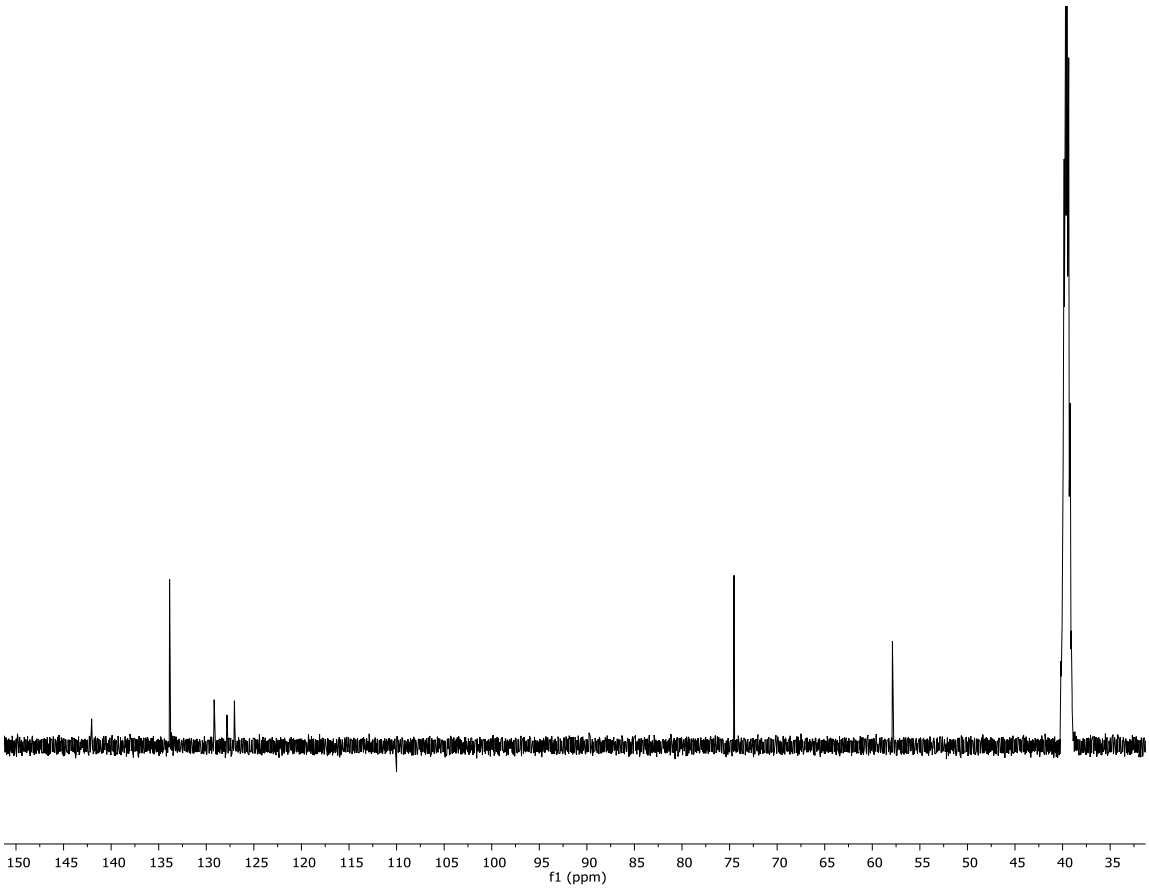


Figure 4.20: Carbon NMR spectra for compound 4.7.

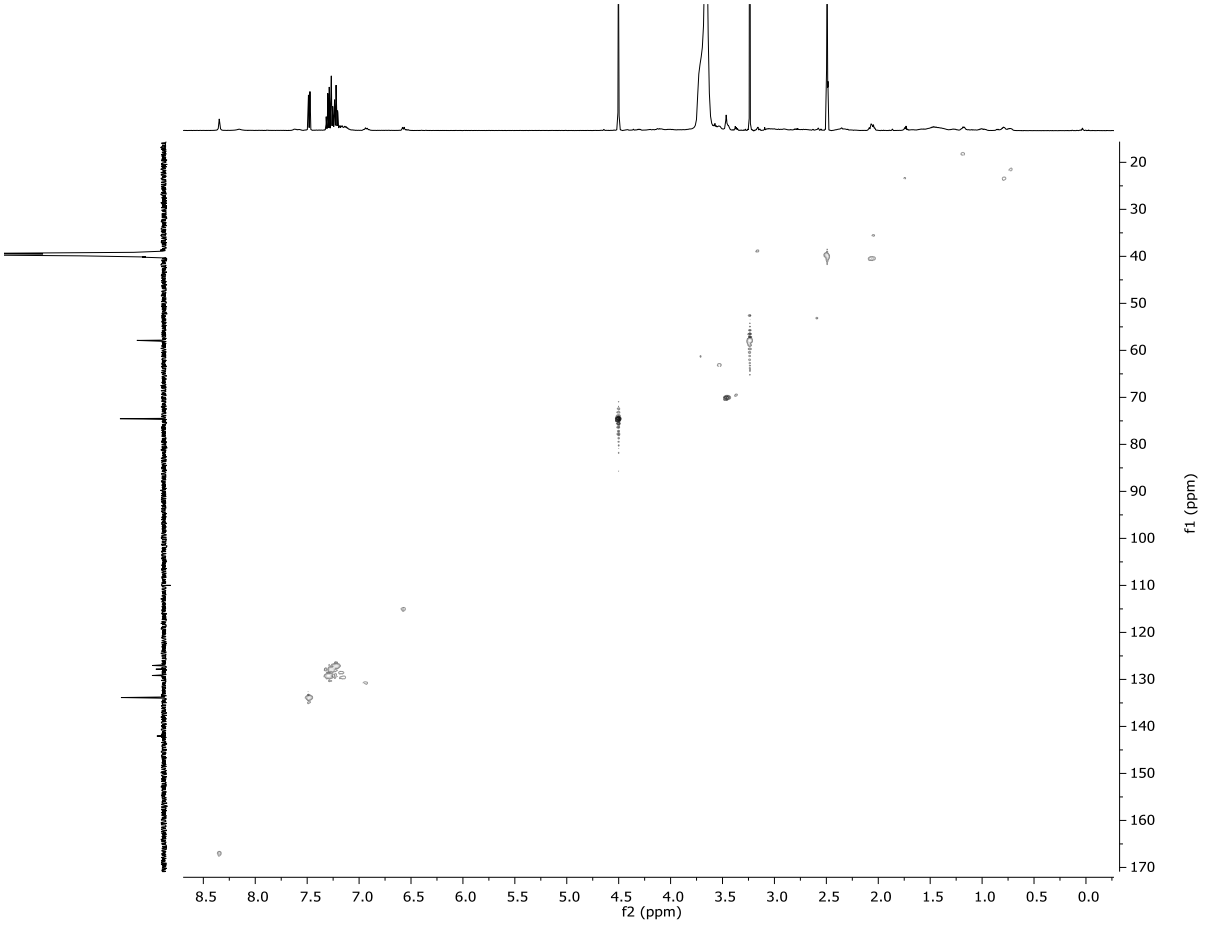
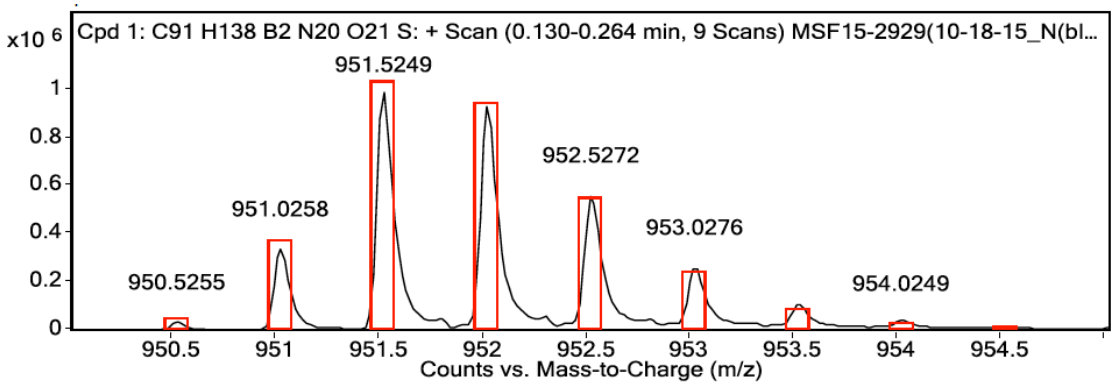
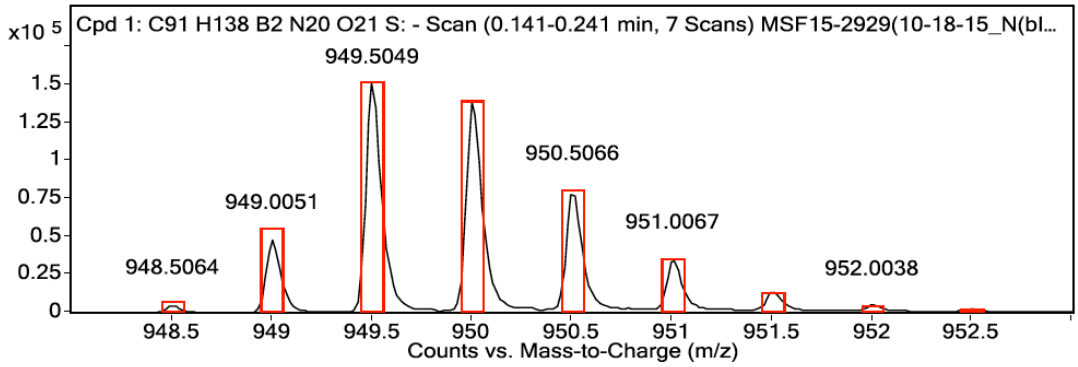


Figure 4.21: C/H correlation NMR for compound 4.7.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
855.49490			1261648.97			
950.52550	950.52350	2	34073.42	C91H138B2N20O21S	(M+2H)+2	-2.04
951.02580	951.02210	2	338403.42	C91H138B2N20O21S	(M+2H)+2	-3.85
951.52490	951.52120	2	987630.58	C91H138B2N20O21S	(M+2H)+2	-3.94
952.02610	952.02210	2	931304.14	C91H138B2N20O21S	(M+2H)+2	-4.2
952.52720	952.52300	2	562050.62	C91H138B2N20O21S	(M+2H)+2	-4.44
953.02760	953.02380	2	261856.43	C91H138B2N20O21S	(M+2H)+2	-3.94
953.52690	953.52470	2	96833.83	C91H138B2N20O21S	(M+2H)+2	-2.38
954.02490	954.02550	2	28395.55	C91H138B2N20O21S	(M+2H)+2	0.65
954.52010	954.52640	2	9538.52	C91H138B2N20O21S	(M+2H)+2	6.63



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
871.49150			617211.03			
948.50640	948.50900	2	4929.35	C91H138B2N20O21S	(M-2H)-2	2.71
949.00510	949.00750	2	47806.59	C91H138B2N20O21S	(M-2H)-2	2.59
949.50490	949.50660	2	151161.6	C91H138B2N20O21S	(M-2H)-2	1.79
950.00610	950.00760	2	139601.2	C91H138B2N20O21S	(M-2H)-2	1.49
950.50660	950.50840	2	79919.56	C91H138B2N20O21S	(M-2H)-2	1.98
951.00670	951.00930	2	34830.69	C91H138B2N20O21S	(M-2H)-2	2.66
951.50610	951.51010	2	12265.04	C91H138B2N20O21S	(M-2H)-2	4.23
952.00380	952.01100	2	4157.1	C91H138B2N20O21S	(M-2H)-2	7.49
952.49640	952.51190	2	1517.84	C91H138B2N20O21S	(M-2H)-2	16.21

Figure 4.22: HRMS data for compound 4.7.

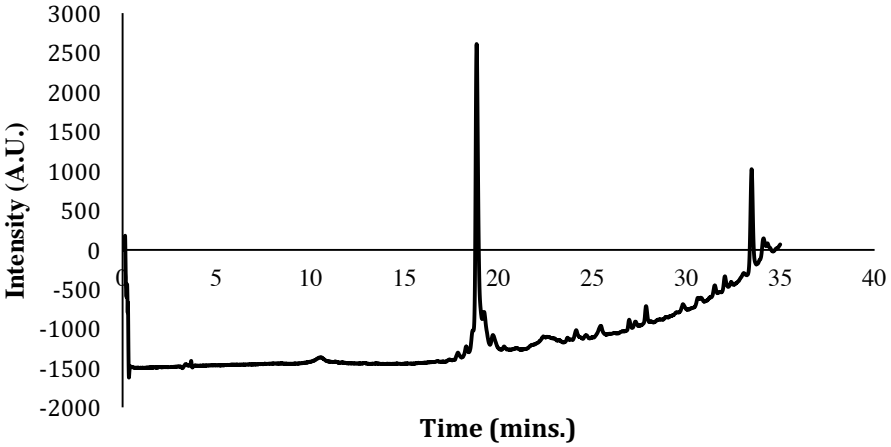
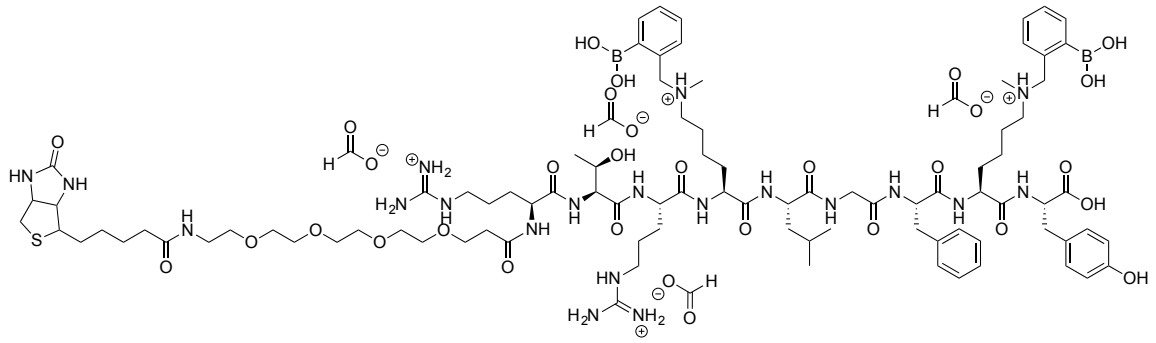


Figure 4.23: Purity check for compound 4.7. Retention time: 17.995 min.



4.8

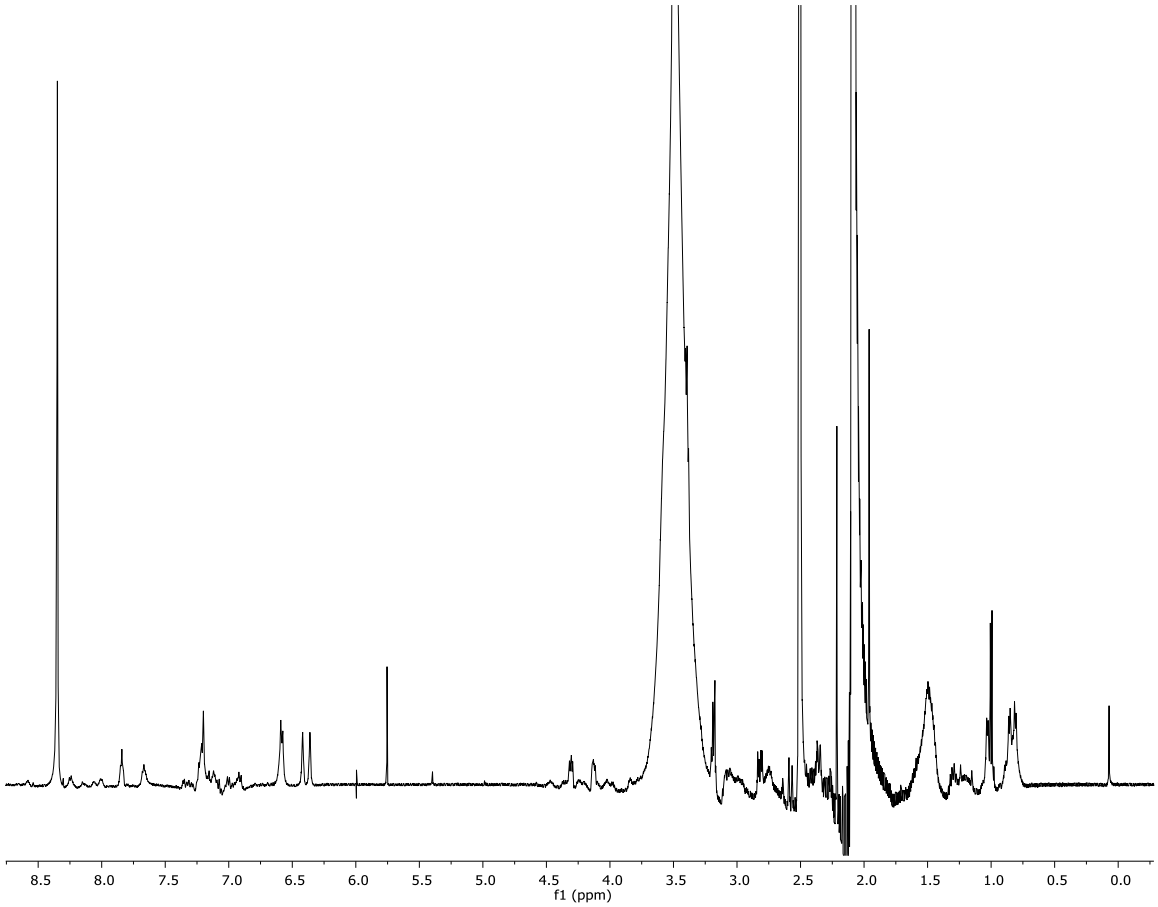


Figure 4.24: Proton NMR spectra for compound 4.8.

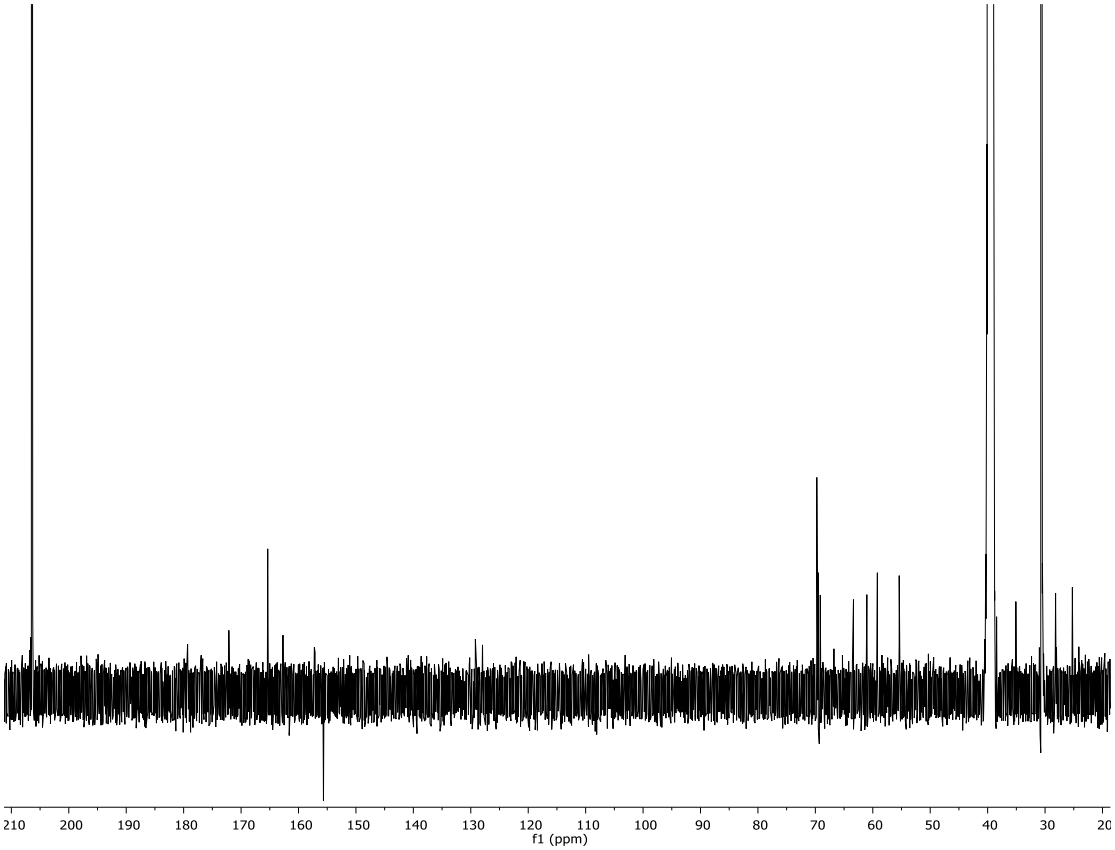
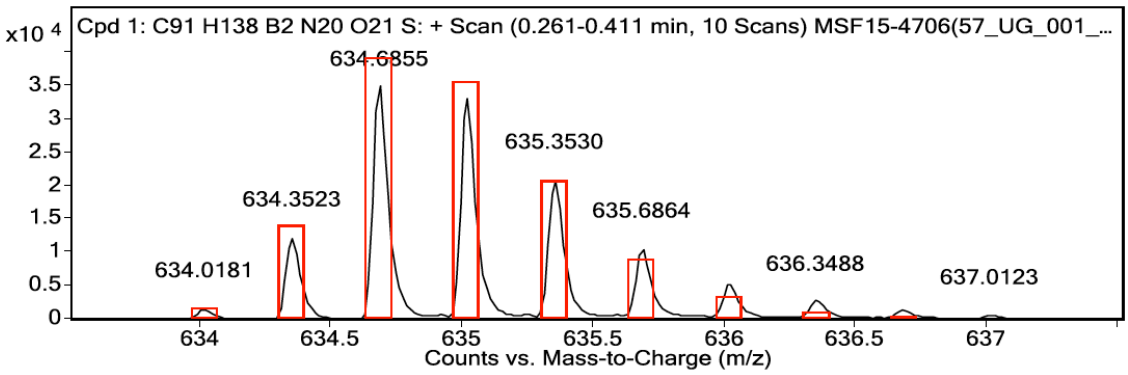


Figure 4.25. Carbon NMR spectra for compound 4.8.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
480.76890			88331.52			
634.01810	634.01810	3	1567.12	C91H138B2N20O21S	(M+3H)+3	0.04
634.35230	634.35050	3	12233.25	C91H138B2N20O21S	(M+3H)+3	-2.86
634.68550	634.68320	3	35162.83	C91H138B2N20O21S	(M+3H)+3	-3.65
635.01900	635.01720	3	33350.68	C91H138B2N20O21S	(M+3H)+3	-2.84
635.35300	635.35110	3	20894.16	C91H138B2N20O21S	(M+3H)+3	-3.03
635.68640	635.68500	3	10532.05	C91H138B2N20O21S	(M+3H)+3	-2.19
636.01780	636.01890	3	5506.16	C91H138B2N20O21S	(M+3H)+3	1.7
636.34880	636.35280	3	2540.84	C91H138B2N20O21S	(M+3H)+3	6.18
636.67980	636.68670	3	1227.04	C91H138B2N20O21S	(M+3H)+3	10.85

Figure 4.26: HRMS data for compound 4.8.

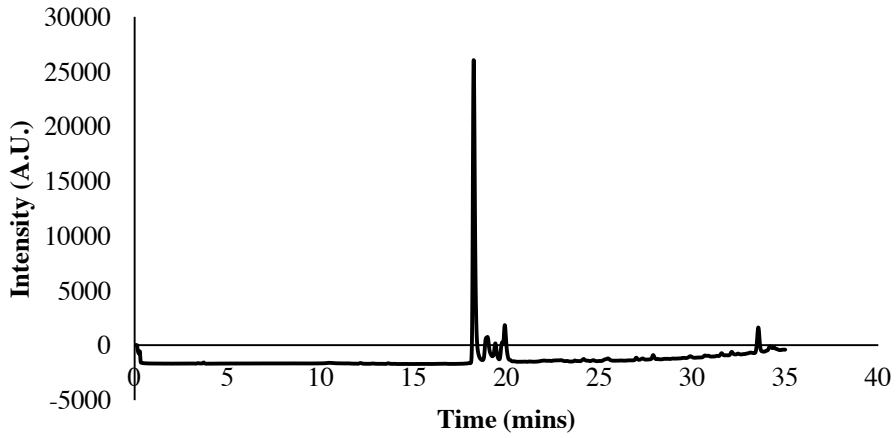
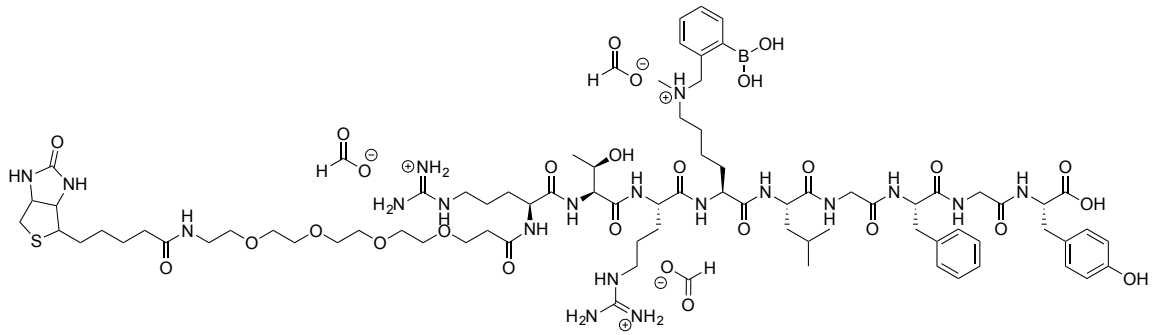


Figure 4.27: Purity check for compound 4.8. Retention time: 18.240 min.



4.9

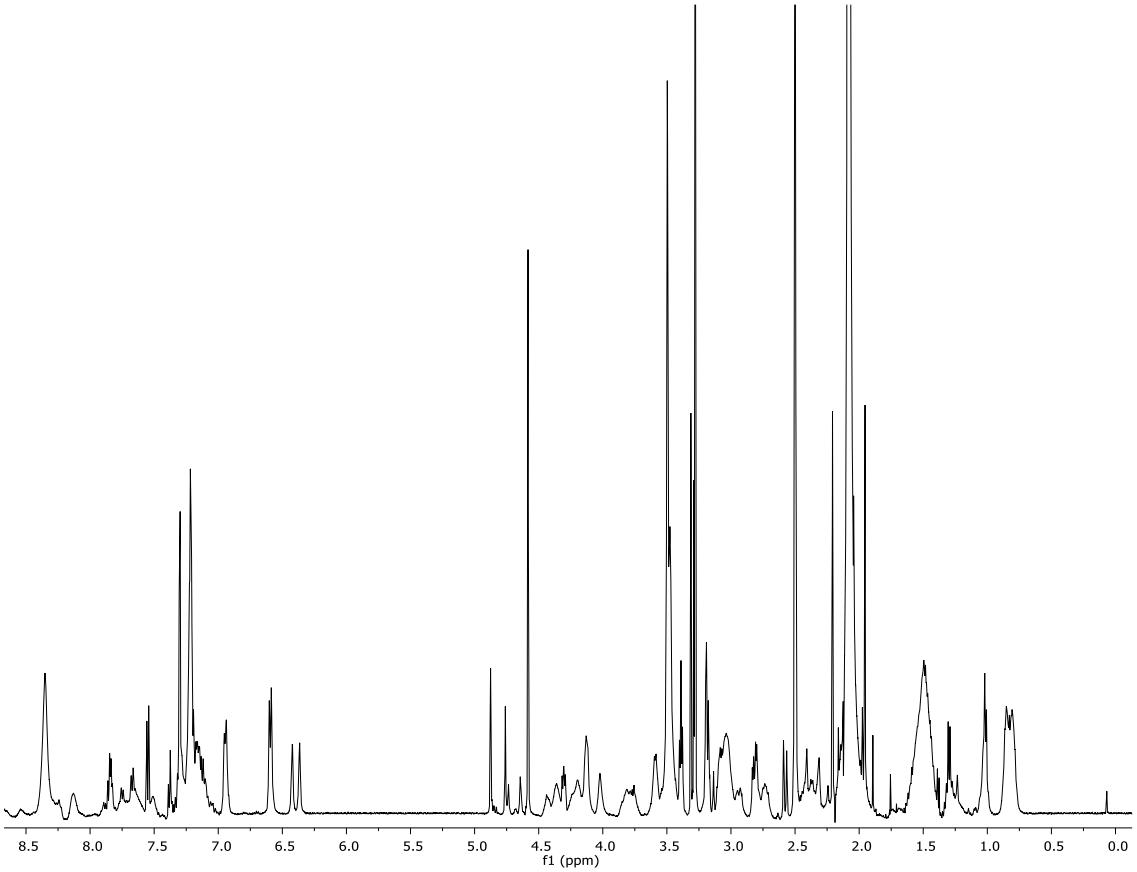


Figure 4.28. Proton NMR spectra for compound 4.9.

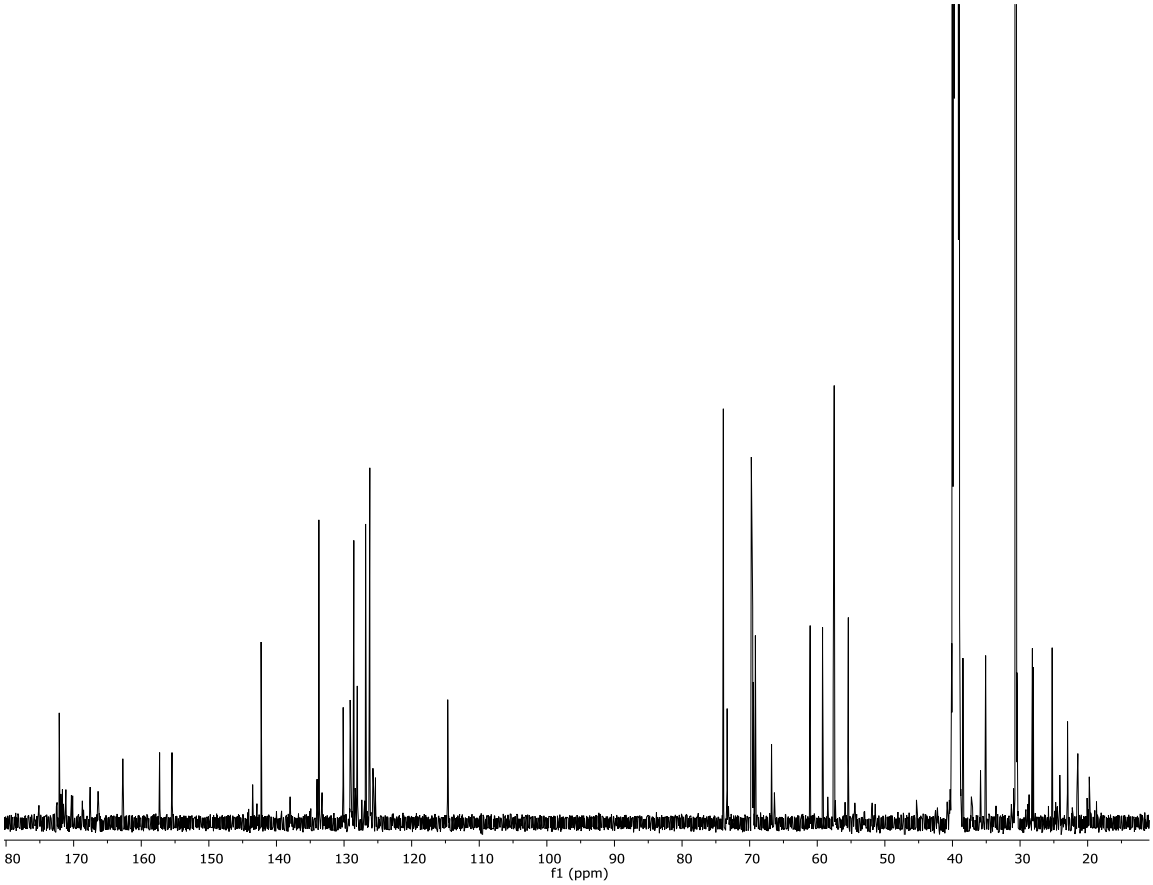
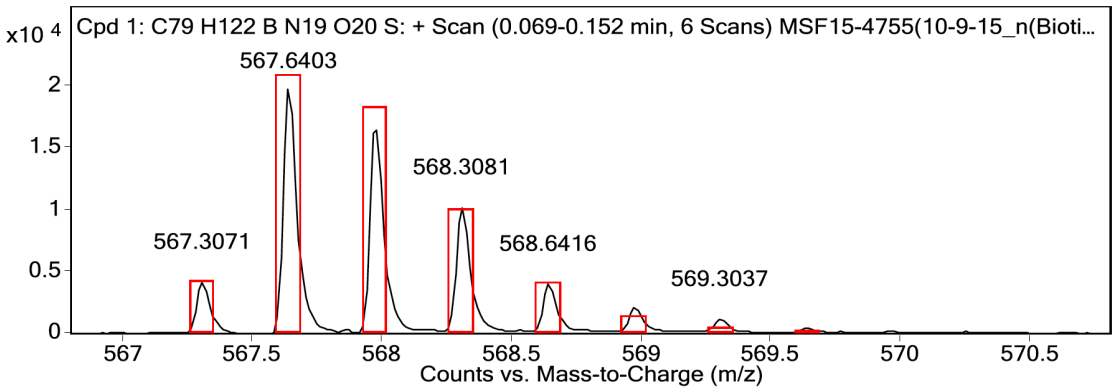
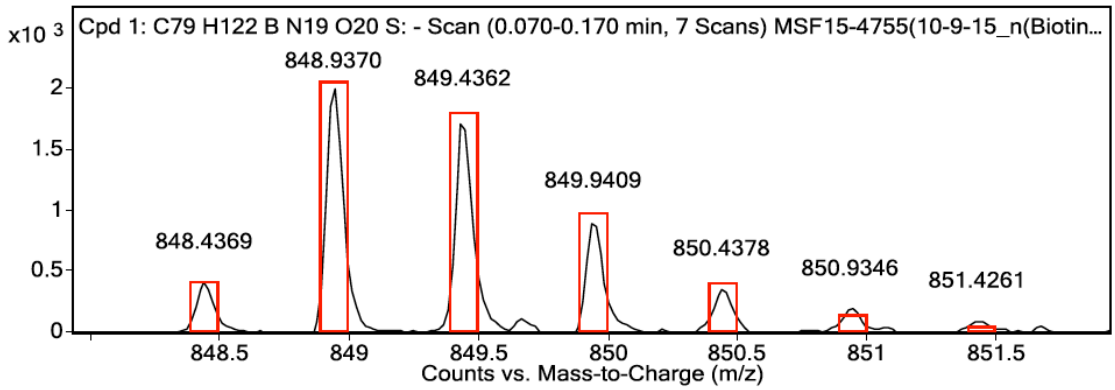


Figure 4.29. Carbon NMR spectra for compound 4.9.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
567.30710	567.30610	3	4174.34	C79H122BN19O20S	(M+3H)+3	-1.78
567.64030	567.63860	3	19997.81	C79H122BN19O20S	(M+3H)+3	-2.95
567.97460	567.97270	3	16917.67	C79H122BN19O20S	(M+3H)+3	-3.29
568.30810	568.30660	3	10229.43	C79H122BN19O20S	(M+3H)+3	-2.6
568.64160	568.64050	3	4013.16	C79H122BN19O20S	(M+3H)+3	-2.02
568.97360	568.97440	3	2116.38	C79H122BN19O20S	(M+3H)+3	1.3
569.30370	569.30830	3	1158.28	C79H122BN19O20S	(M+3H)+3	8.06
569.63940	569.64220	3	493.28	C79H122BN19O20S	(M+3H)+3	4.88



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
812.41500			2168.39			
848.43690	848.44090	2	420.19	C79H122BN19O20S	(M-2H)-2	4.71
848.93700	848.93970	2	2022.27	C79H122BN19O20S	(M-2H)-2	3.21
849.43620	849.44090	2	1758.98	C79H122BN19O20S	(M-2H)-2	5.55
849.94090	849.94180	2	922.05	C79H122BN19O20S	(M-2H)-2	0.97
850.43780	850.44260	2	364.24	C79H122BN19O20S	(M-2H)-2	5.6
850.93460	850.94340	2	200.64	C79H122BN19O20S	(M-2H)-2	10.26
851.42610	851.44420	2	101.37	C79H122BN19O20S	(M-2H)-2	21.24

Figure 4.30: HRMS data for compound 4.9.

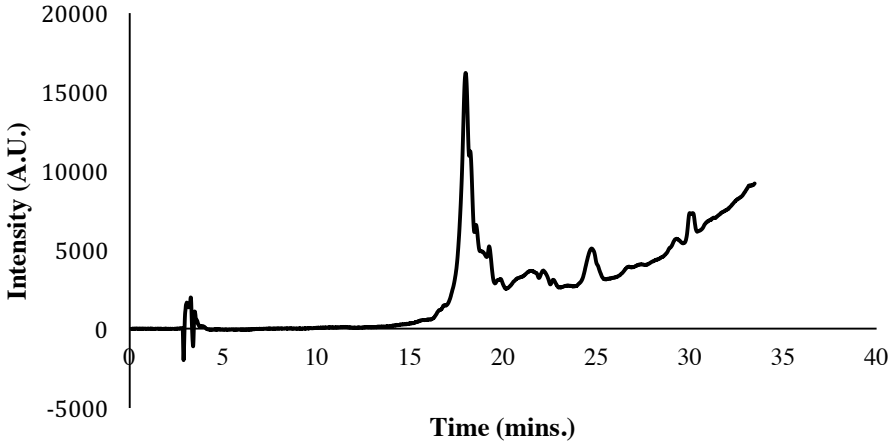
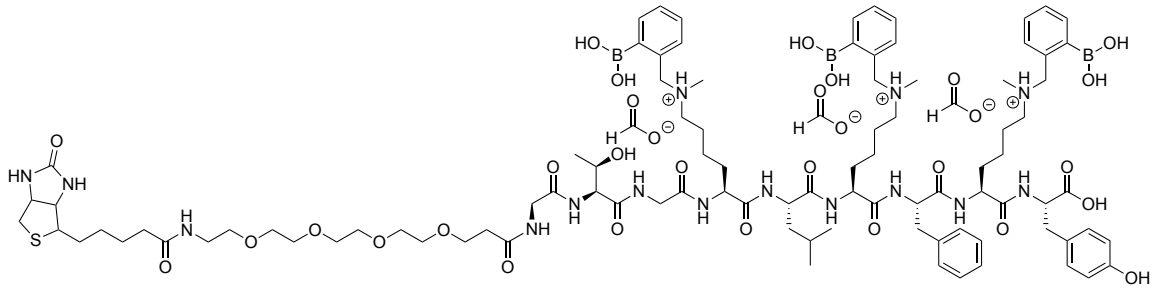


Figure 4.31: Purity check for compound 4.9. Retention time: 17.995 min.



4.10

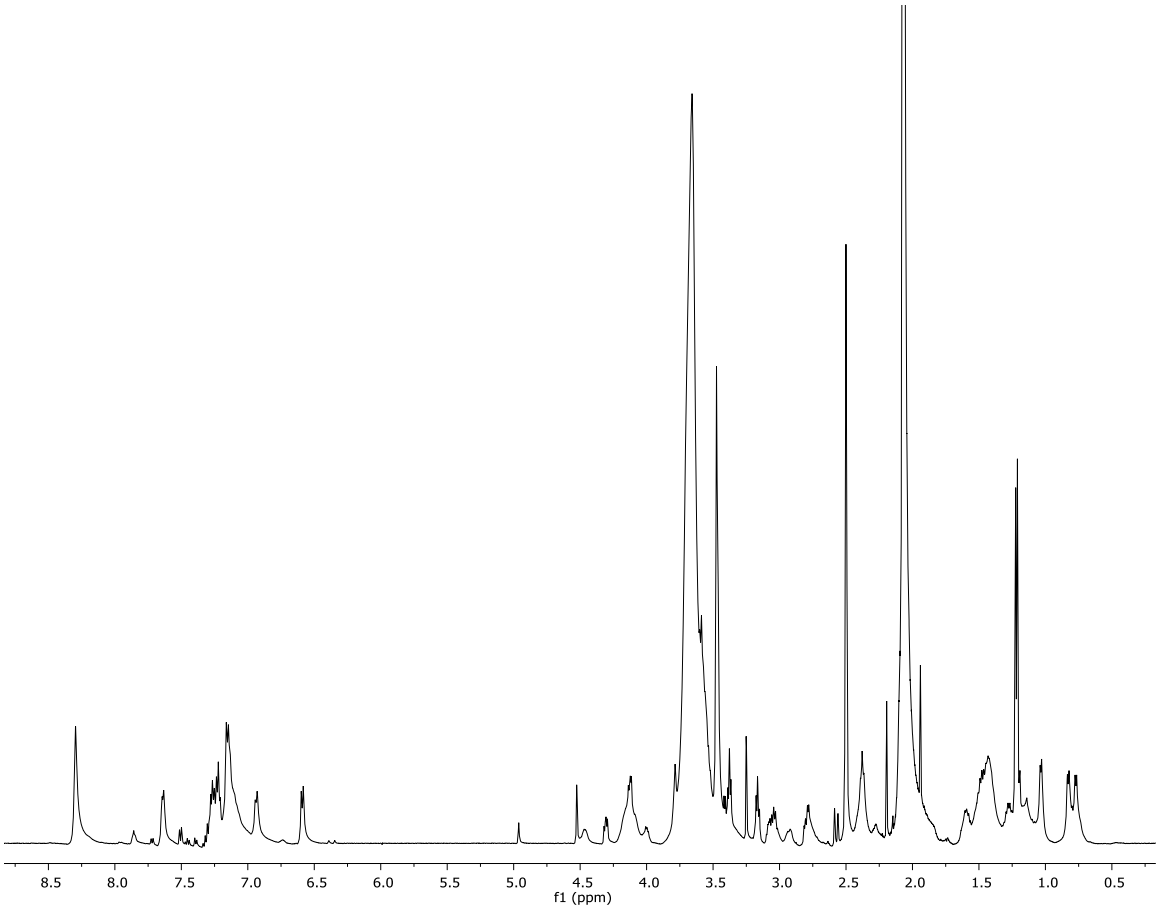


Figure 4.32: Proton NMR spectra for compound 4.10.

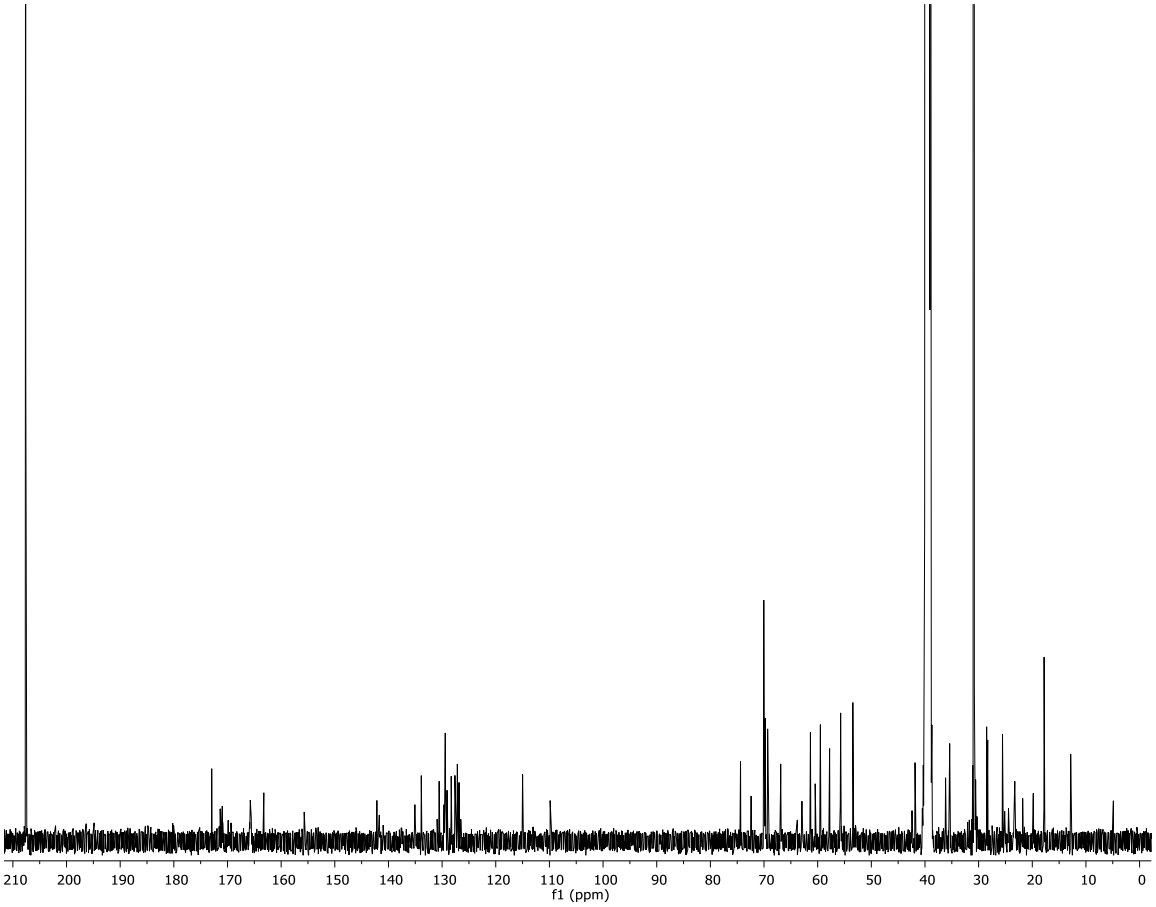
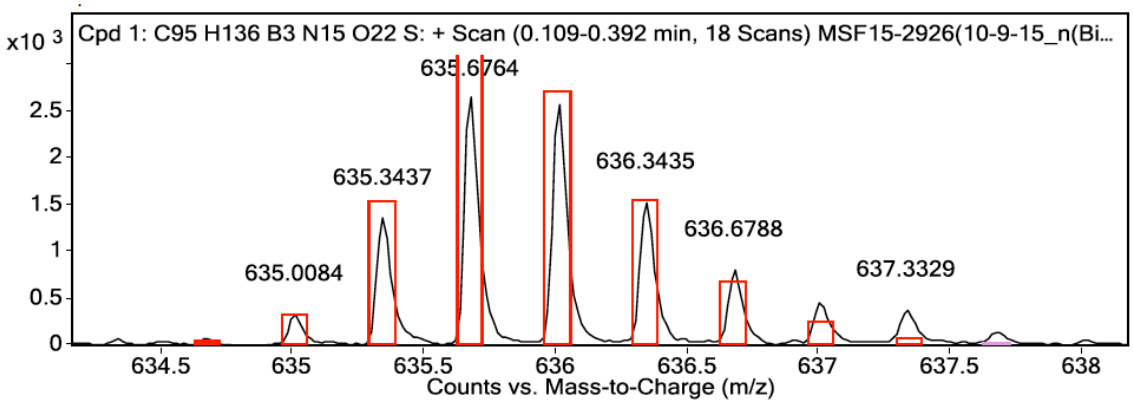
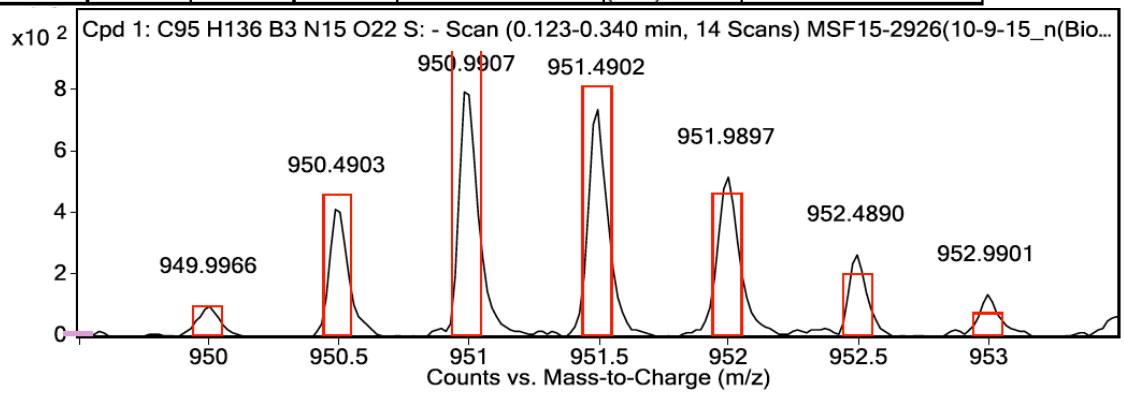


Figure 4.33: Carbon NMR spectra for compound 4.10.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
130.15890			28339.29			
634.67140	634.67700	3	83.41	C95H136B3N15O22S	(M+3H)+3	8.93
635.00840	635.00940	3	334.79	C95H136B3N15O22S	(M+3H)+3	1.44
635.34370	635.34180	3	1359.94	C95H136B3N15O22S	(M+3H)+3	-2.91
635.67640	635.67470	3	2658.66	C95H136B3N15O22S	(M+3H)+3	-2.81
636.01060	636.00850	3	2575.64	C95H136B3N15O22S	(M+3H)+3	-3.27
636.34350	636.34240	3	1532.21	C95H136B3N15O22S	(M+3H)+3	-1.71
636.67880	636.67630	3	807.59	C95H136B3N15O22S	(M+3H)+3	-3.9
637.00430	637.01020	3	456.44	C95H136B3N15O22S	(M+3H)+3	9.3
637.33290	637.34410	3	379.1	C95H136B3N15O22S	(M+3H)+3	17.66



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
949.99660	949.99580	2	99.49	C95H136B3N15O22S	(M-2H)-2	-0.8
950.49030	950.49450	2	427.93	C95H136B3N15O22S	(M-2H)-2	4.48
950.99070	950.99380	2	829.74	C95H136B3N15O22S	(M-2H)-2	3.26
951.49020	951.49460	2	746.56	C95H136B3N15O22S	(M-2H)-2	4.65
951.98970	951.99540	2	520.63	C95H136B3N15O22S	(M-2H)-2	5.99
952.48900	952.49630	2	268.36	C95H136B3N15O22S	(M-2H)-2	7.64
952.99010	952.99710	2	138.23	C95H136B3N15O22S	(M-2H)-2	7.33
1204.26500			1237.06			

Figure 4.34: HRMS data for compound 4.10.

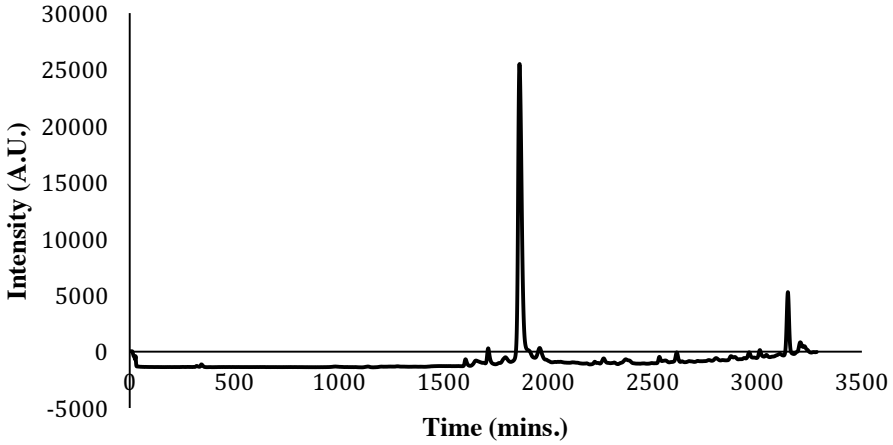
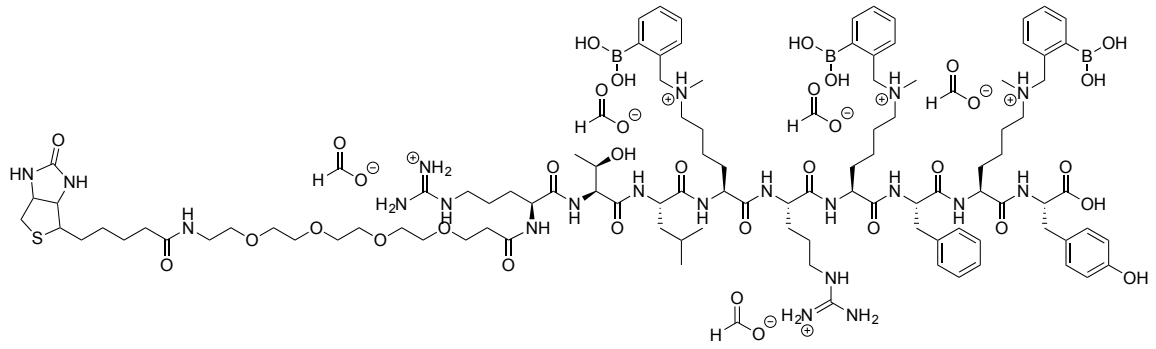


Figure 4.35: Purity check for compound 4.10. Retention time: 19.861 min.



4.11

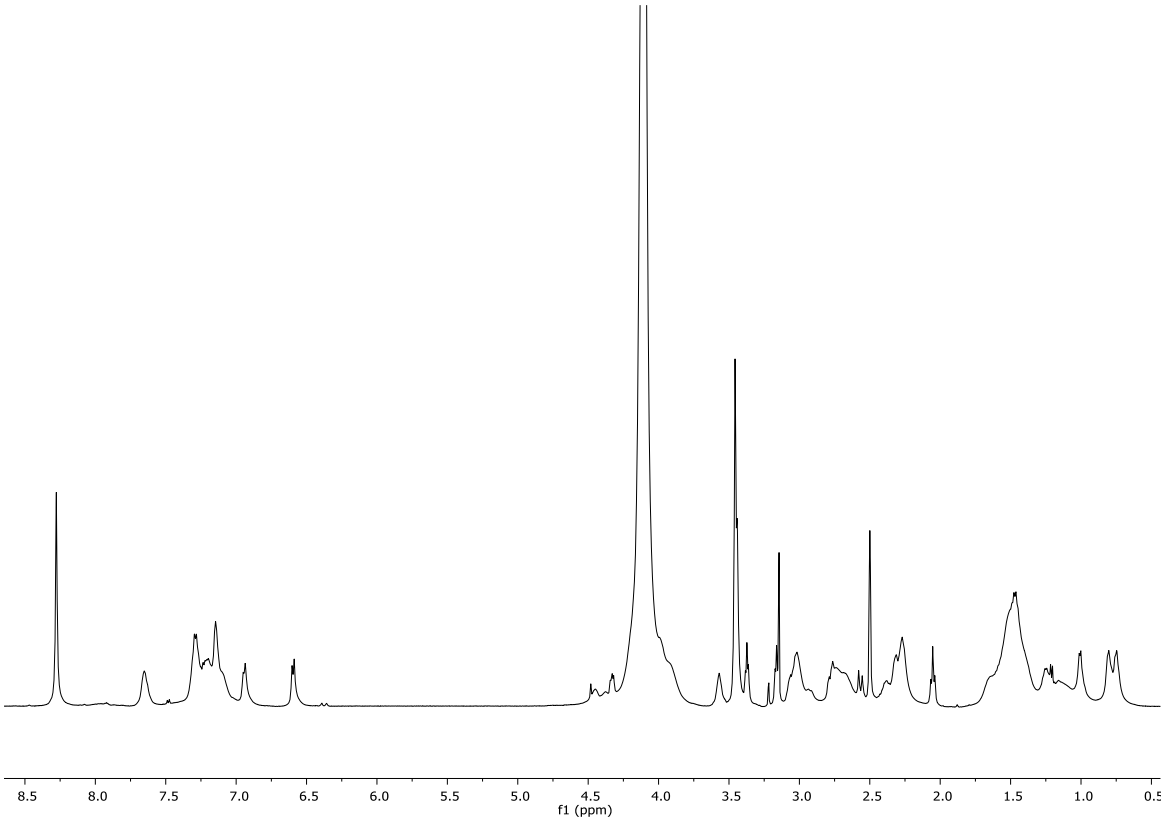


Figure 4.36: Proton NMR spectra for compound 4.11.

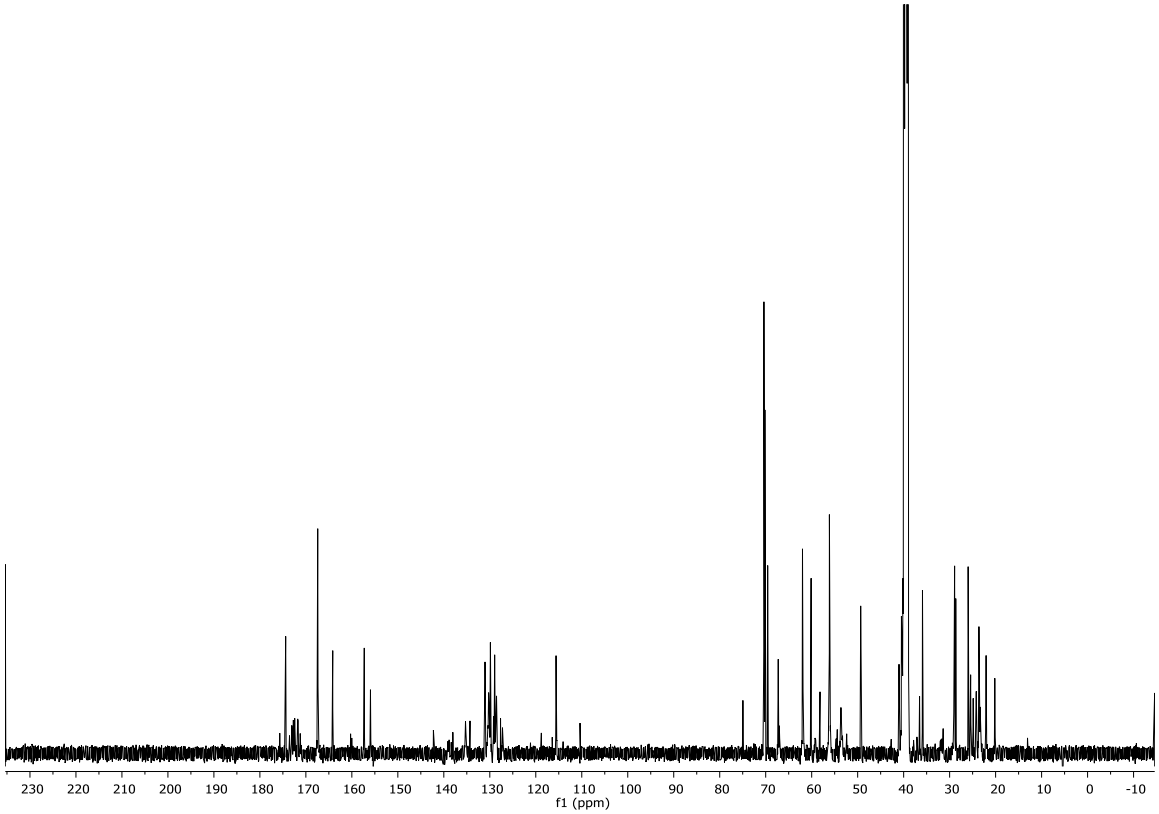


Figure 4.37: Carbon NMR spectra for compound 4.11.

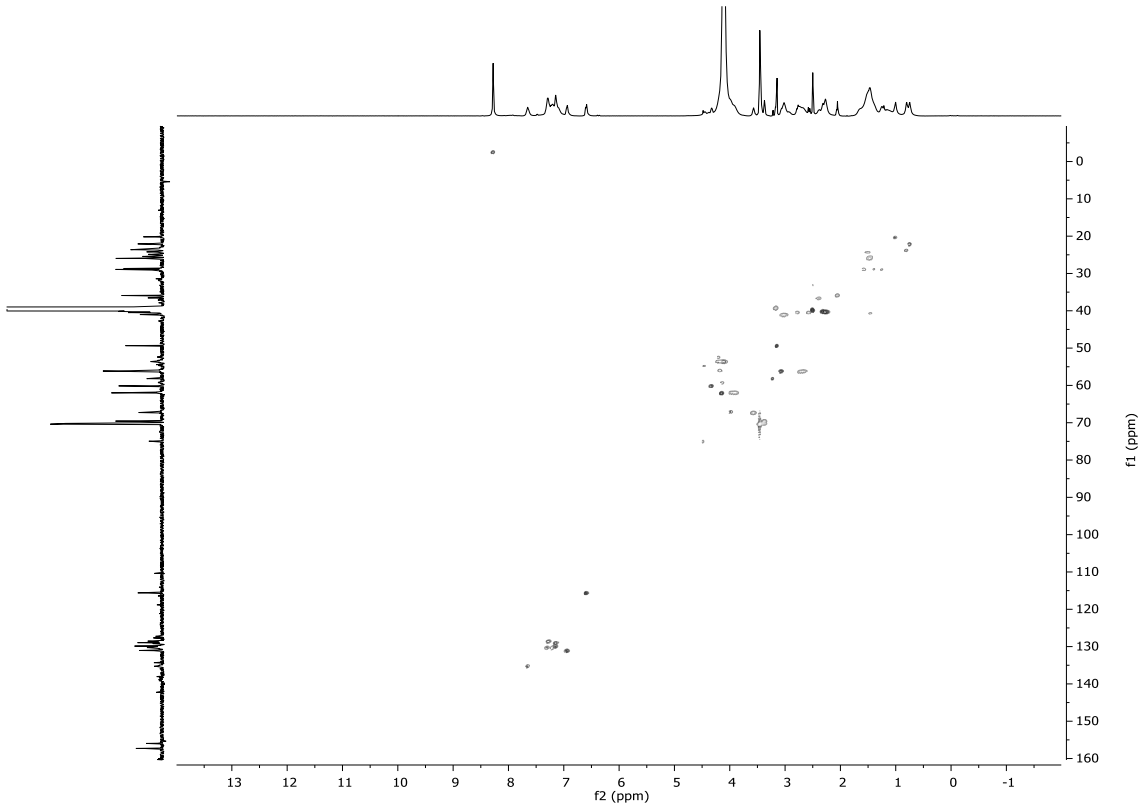
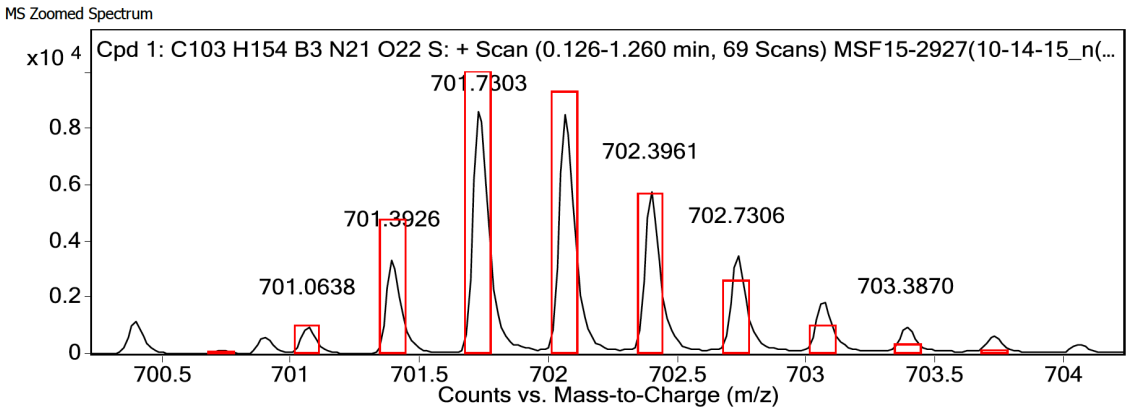
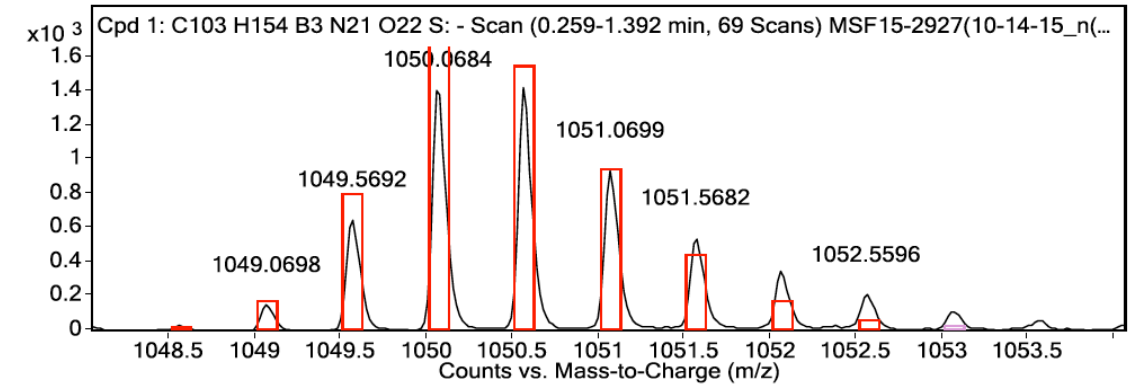


Figure 4.38: C/H correlation NMR for compound 4.11.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
130.15910			34928.27			
700.73060	700.73010	3	175.78	C103H154B3N21O22S	(M+3H)+3	-0.61
701.06380	701.06250	3	969.65	C103H154B3N21O22S	(M+3H)+3	-1.85
701.39260	701.39500	3	3391.49	C103H154B3N21O22S	(M+3H)+3	3.39
701.73030	701.72780	3	8844.23	C103H154B3N21O22S	(M+3H)+3	-3.49
702.06400	702.06160	3	8637.74	C103H154B3N21O22S	(M+3H)+3	-3.4
702.39610	702.39550	3	5805.99	C103H154B3N21O22S	(M+3H)+3	-0.75
702.73060	702.72940	3	3515.3	C103H154B3N21O22S	(M+3H)+3	-1.62
703.06250	703.06330	3	1875.76	C103H154B3N21O22S	(M+3H)+3	1.16
703.38700	703.39730	3	999.65	C103H154B3N21O22S	(M+3H)+3	14.53



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
1048.56380	1048.57700	2	34	C103H154B3N21O22S	(M-2H)-2	12.66
1049.06980	1049.07550	2	149.97	C103H154B3N21O22S	(M-2H)-2	5.43
1049.56920	1049.57420	2	648.83	C103H154B3N21O22S	(M-2H)-2	4.77
1050.06840	1050.07350	2	1438.84	C103H154B3N21O22S	(M-2H)-2	4.91
1050.56820	1050.57430	2	1428.76	C103H154B3N21O22S	(M-2H)-2	5.82
1051.06990	1051.07510	2	936.71	C103H154B3N21O22S	(M-2H)-2	5
1051.56820	1051.57600	2	542.3	C103H154B3N21O22S	(M-2H)-2	7.4
1052.06280	1052.07680	2	352.52	C103H154B3N21O22S	(M-2H)-2	13.34
1052.55960	1052.57770	2	209.53	C103H154B3N21O22S	(M-2H)-2	17.16

Figure 4.39: HRMS data for compound 4.11.

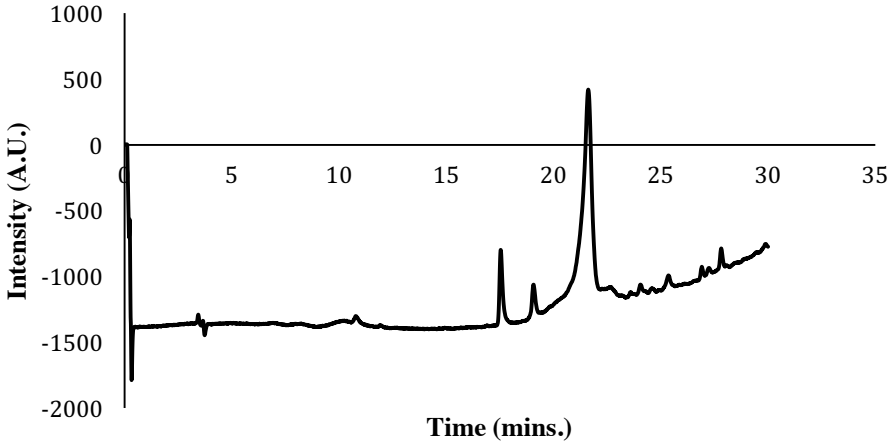
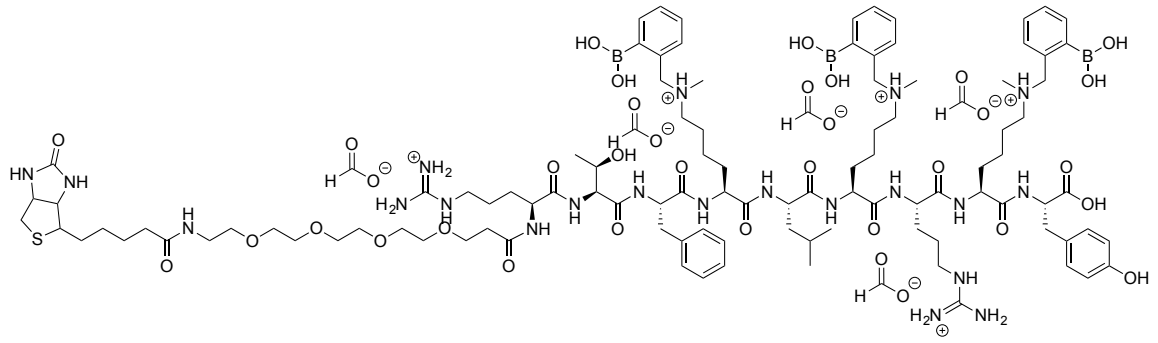


Figure 4.40: Purity check for compound 4.11. Retention time: 21.621 min.



4.12

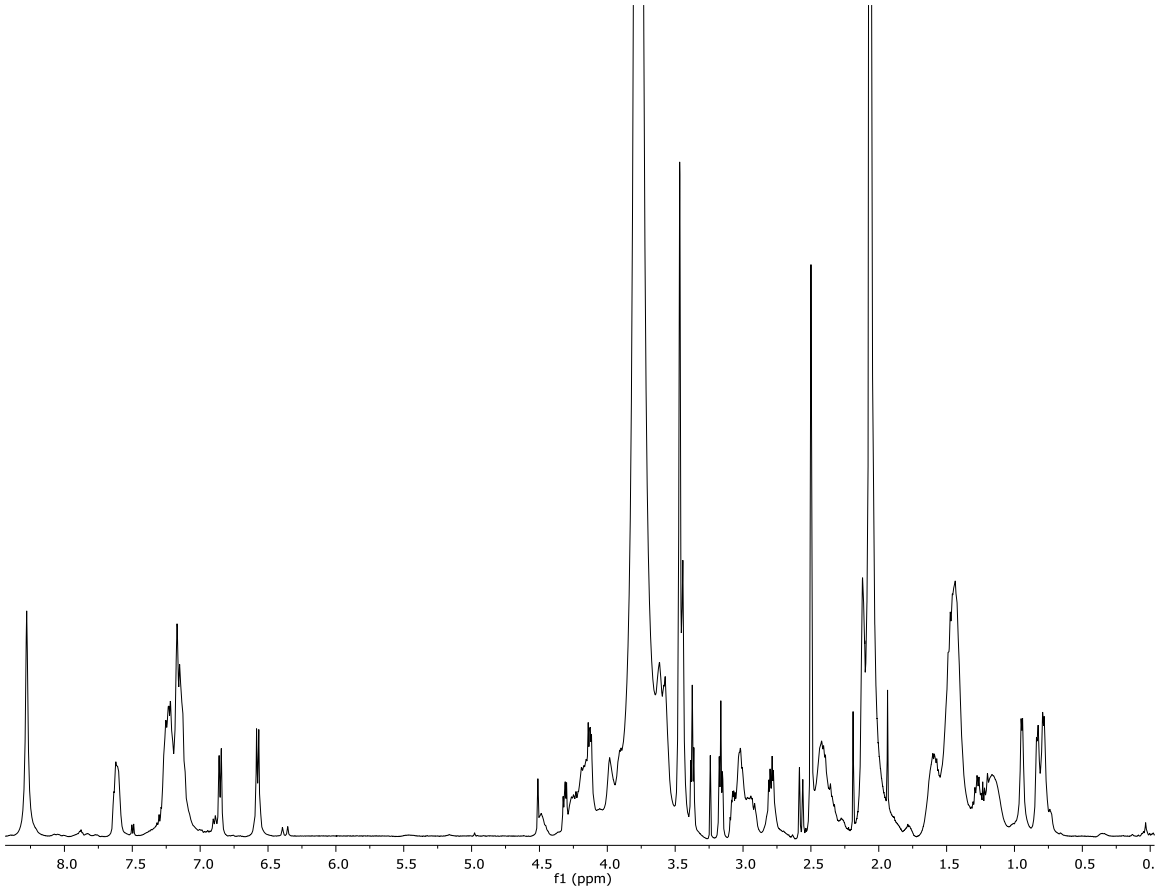
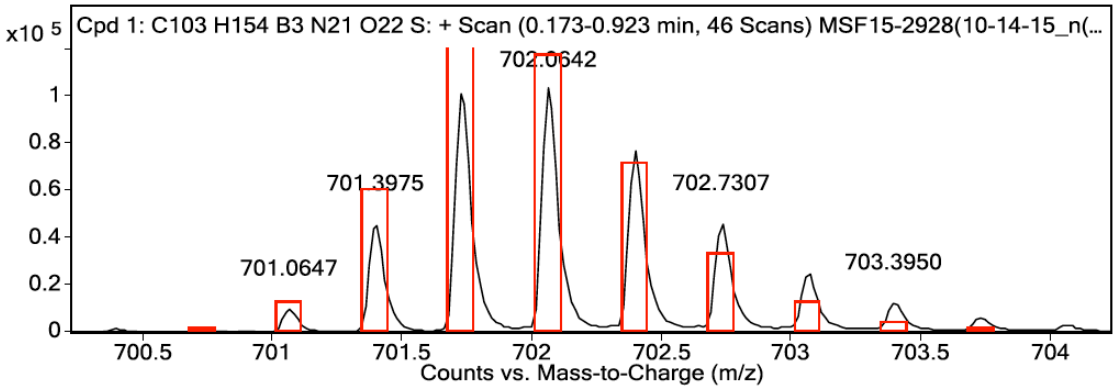
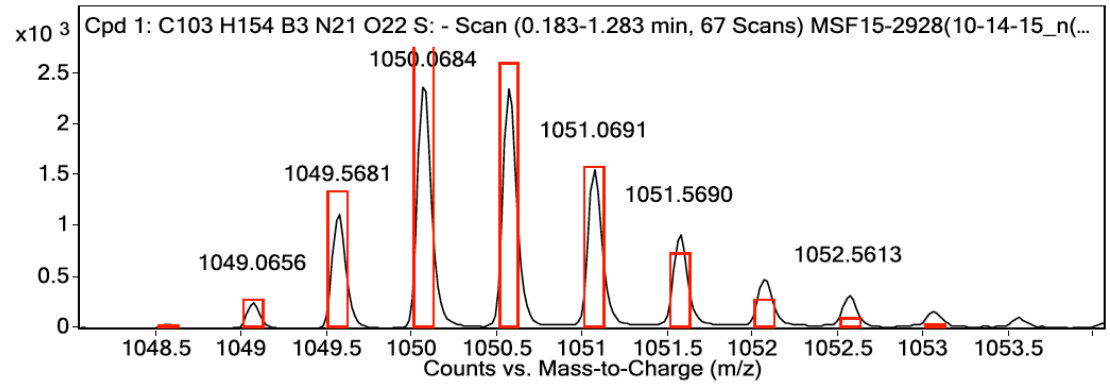


Figure 4.41: Proton NMR spectra for compound 4.12.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
700.73090	700.73010	3	1945.73	C103H154B3N21O22S	(M+3H)+3	-1.08
701.06470	701.06250	3	10041.99	C103H154B3N21O22S	(M+3H)+3	-3.2
701.39750	701.39500	3	46195.99	C103H154B3N21O22S	(M+3H)+3	-3.62
701.73050	701.72780	3	103651.84	C103H154B3N21O22S	(M+3H)+3	-3.77
702.06420	702.06160	3	104893.67	C103H154B3N21O22S	(M+3H)+3	-3.57
702.39750	702.39550	3	77436.78	C103H154B3N21O22S	(M+3H)+3	-2.82
702.73070	702.72940	3	46334.16	C103H154B3N21O22S	(M+3H)+3	-1.82
703.06320	703.06330	3	25027.63	C103H154B3N21O22S	(M+3H)+3	0.16
703.39500	703.39730	3	12624.02	C103H154B3N21O22S	(M+3H)+3	3.28
703.72690	703.73120	3	6397.87	C103H154B3N21O22S	(M+3H)+3	6.11



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
313.14250			3107.64			
1048.56120	1048.57700	2	42.13	C103H154B3N21O22S	(M-2H)-2	15.13
1049.06560	1049.07550	2	251.68	C103H154B3N21O22S	(M-2H)-2	9.42
1049.56810	1049.57420	2	1127.53	C103H154B3N21O22S	(M-2H)-2	5.87
1050.06840	1050.07350	2	2432.65	C103H154B3N21O22S	(M-2H)-2	4.92
1050.56910	1050.57430	2	2367.85	C103H154B3N21O22S	(M-2H)-2	4.95
1051.06910	1051.07510	2	1556.05	C103H154B3N21O22S	(M-2H)-2	5.71
1051.56900	1051.57600	2	922.45	C103H154B3N21O22S	(M-2H)-2	6.61
1052.06580	1052.07680	2	486.66	C103H154B3N21O22S	(M-2H)-2	10.44
1052.56130	1052.57770	2	331.41	C103H154B3N21O22S	(M-2H)-2	15.62

Figure 4.42: HRMS for compound 4.12.

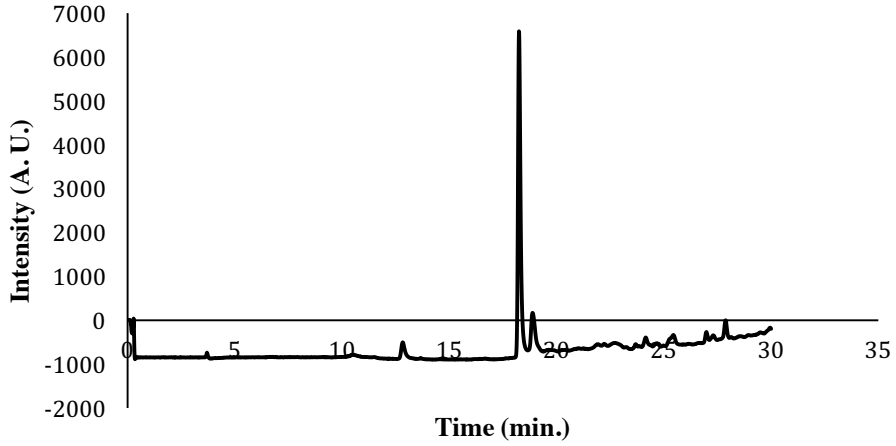


Figure 4.43: Purity check for compound 4.12. Retention time: 18.251 min.

4.7 References

1. S. Abdellaoui, B. C. Corgier, C. A. Mandon, B. Doumeche, C. A. Marquette and L. J. Blum, *Electroanal*, 2013, **25**.
2. R. Amin and S. A. Elfeky, *Spectrochim Acta A*, 2013, **108**.
3. S. Arimori, S. Ushiroda, L. M. Peter, A. T. A. Jenkins and T. D. James, *Chem. Commun.*, 2002, **20**.
4. J. B. Crumpton, W. Zhang and W. L. Santos, *Anal. Chem.*, 2011, **83**, 3548-3554.
5. W. W. Bachovchin, W. Y. L. Wong, S. Farr-Jones, A. B. Shenvi and C. A. Kettner, *Biochemistry (Mosc.)*, 1988, **27**, 7689-7697.
6. W. L. A. Brooks and B. S. Sumerlin, *Chem. Rev.*, 2016, **116**, 1375-1397.
7. J. N. Cambre and B. S. Sumerlin, *Polymer*, 2011, **52**, 4631-4643.
8. T. M. El Dine, J. Rouden and J. Blanchet, *Chem. Commun.*, 2015, **51**, 16084-16087.
9. T. D. James, K. R. A. S. Sandanayake and S. Shinkai, *Angew Chem Int Edit*, 1994, **33**.
10. S. Kotha, K. Lahiri and D. Kashinath, *Tetrahedron*, 2002, **58**.
11. K. Lacina and P. Skládal, *Electrochim Acta*, 2011, **56**.
12. K. Lacina, P. Skládal and T. D. James, *Chem. Cent. J.*, 2014, **8**, 1-17.
13. J. Langen, U. Fischer, M. Cavalar, C. Coetzee, P. Wegmann-Herr and H.-G. Schmarr, *Anal. Bioanal. Chem.*, 2016, **408**, 2425-2439.
14. J. Li, A. S. Grillo and M. D. Burke, *Accounts of Chemical Research*, 2015, **48**, 2297-2307.
15. M. Li, W. Zhu, F. Marken and T. D. James, *Chem. Commun.*, 2015, **51**, 14562-14573.
16. J. X. Qiao and P. Y. S. Lam, *Synthesis*, 2011, **6**.

17. L. Rocard, A. Berezin, F. De Leo and D. Bonifazi, *Angew. Chem. Int. Ed.*, 2015, **54**, 15739-15743.
18. G. Springsteen and B. Wang, *Chem. Commun.*, 2001, **17**.
19. X. Sun, W. Zhai, J. S. Fossey and T. D. James, *Chem. Commun.*, 2016, **52**, 3456-3469.
20. Q. Zhang, N. Tang, J. W. C. Brock, H. M. Mottaz, J. M. Ames, J. W. Baynes, R. D. Smith and T. O. Metz, *J. Proteome Res.*, 2007, **6**, 2323-2330.
21. W. Zhang, D. I. Bryson, J. B. Crumpton, J. Wynn and W. L. Santos, *Chem. Commun.*, 2013, **49**, 2436-2438.
22. X. Sun and T. D. James, *Chem. Rev.*, 2015, **115**, 8001-8037.
23. S.-H. Chung, T.-J. Lin, Q.-Y. Hu, C.-H. Tsai and P.-S. Pan, *Molecules*, 2013, **18**, 12346.
24. J. J. Deadman, S. Elgendy, C. A. Goodwin, D. Green, J. A. Baban, G. Patel, E. Skordalakes, N. Chino and G. Claeson, *J. Med. Chem.*, 1995, **38**, 1511-1522.
25. P. J. Duggan and D. A. Offermann, *Aust. J. Chem.*, 2007, **60**, 829-834.
26. N. Y. Edwards, T. W. Sager, J. T. McDevitt and E. V. Anslyn, *J. Am. Chem. Soc.*, 2007, **129**, 13575-13583.
27. G. A. Ellis, M. J. Palte and R. T. Raines, *J. Am. Chem. Soc.*, 2012, **134**, 3631-3634.
28. O. V. Gozhina, J. S. Svendsen and T. Lejon, *J. Pep. Sci.*, 2014, **20**, 20-24.
29. M. Kita, J. Yamamoto, T. Morisaki, C. Komiya, T. Inokuma, L. Miyamoto, K. Tsuchiya, A. Shigenaga and A. Otaka, *Tet. Lett.*, 2015, **56**, 4228-4231.
30. C.-H. Tsai, C.-H. Lin, C.-T. Hsieh, C.-C. Cai, T.-J. Lin, P.-Y. Liu, M.-H. Lin, M.-J. Wu, C.-C. Fu, Y.-C. Wu, F.-R. Chang and P.-S. Pan, *Res. Chem. Intermed.*, 2014, **40**, 2187-2198.
31. K. L. Bicker, J. Sun, J. J. Lavigne and P. R. Thompson, *ACS Comb. Sci.*, 2011, **13**, 232-243.
32. B. E. Collins and E. V. Anslyn, *Chem. European J.*, 2007, **13**, 4700-4708.
33. B. E. Collins, S. Sorey, A. E. Hargrove, S. H. Shabbir, V. M. Lynch and E. V. Anslyn, *J. Org. Chem.*, 2009, **74**, 4055-4060.
34. D. G. Hall, in *Boronic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, 2006, pp. 1-99.
35. J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem. .Asian J.*, 2009, **4**, 630-640.
36. N. Fomina, C. McFearin, M. Sermsakdi, O. Edigin and A. Almutairi, *J. Am. Chem. Soc.*, 2010, **132**, 9540-9542.
37. V. Castro, H. Rodríguez and F. Albericio, *ACS Comb. Sci.*, 2016, **18**, 1-14.

Chapter 5: The Synthesis of Bipyridine Peptides via Alkylation with 5- and 6-(bromomethyl)-2,2'-bipyridine for Metal-based Cyclization Studies in the Presence of Hydrazone Forming Peptides

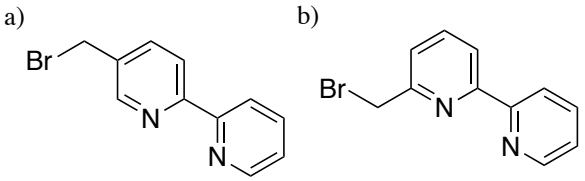
5.1 INTRODUCTION

5.1.1 Metals and Ligands in Forming Higher-Order Organic Structures

Self-assembly of small organic molecules or oligomers into large complex structures resemble biomolecule folding in nature. Thus, when chemists observe such processes in synthetic systems, these higher order structures are called "biomimetic." Common metal ligands in coordination chemistry are studied for folding into structures. These ligands garner interest because of the ability to control their arrangement in space by a metal. If conformational changes are made at will, using externally applied stimuli, parallels to metal based protein structuring and aggregation in nature can be made.[1]

5.1.2 Facilitating Incorporation of Ligands into Peptides

As described in Chapters 2-4, the chemistries presented so far aimed at the selective modification of amino acids for protein sequencing can also be applied.[2] Alkylation of secondary amine residues was the method chosen for incorporation of boronic acids.[3] Use of similar methodology to modify peptides with 2-2'-bipyridines is presented here. Including N-methyl Lysine via solid-phase peptide synthesis (SPPS) remained the manner in which secondary amines were introduced. Brominated compounds 5.1 and 5.2 were synthesized for this purpose. To our knowledge, alkylation is not a reported method for functionalizing peptides with bipyridines. However, it is for the synthesis of small molecule ligands.



5.1 5.2

Figure 5.1: Brominated bipyridines. a) 5-(bromomethyl)-2,2'-bipyridine b) 6-(bromomethyl)-2,2'-bipyridine

5.1.3 Bipyridine Ligands

The use of one pyridine ring, in conjunction with complimentary natural/ metal binder, has led to interesting folding properties. Thus, increasing the number of pyridines for formation of different ligands is part of the reason terpyridines and other multi-dentate, heterocyclic aromatic structures are also employed. Bipyridine is one of the most widely used metal ligands. Synthesis of small molecules, polymers, and peptides containing this functional group has been reported.[4,5] These moieties facilitate helical formation of organic compounds, act as materials in photovoltaic/electronic devices, serve as achiral molecular sensors, and are applied in luminescent devices.[4] Peptide chemists have found their utility quite desirable and amino acid building blocks with this moiety have been made.[6] The need to synthesize building blocks adds to the onerous task of synthesizing large amounts of starting material, mostly which is discarded during peptide synthesis. Nevertheless, methods to incorporate bipyridines has been extensively studied and analyzed for their peptide folding properties.[7] These moieties are regarded as an integral part in the chemist's effort to materialize protein-like properties from unnatural functionalities.[8]

5.1.4 Relevance for Unnatural Metal-Based Peptides

Proteins and peptides are polymers that nature uses as a scaffold for the modular synthesis of macromolecules numbering conservatively ~ 20, 000 proteins before post-translation modifications (PTMs).[9] These biomolecules execute a diverse set of functions

inside the body and are composed of a relatively simple set of organic side chain moieties and functionalities. Nearly one-third of proteins have metal coordination sites to execute regulatory or enzymatic processes.[7] Therefore, peptide chemists have tried to replicate metal ligand coordination with both naturally occurring residues and non-canonical metal binders. Bipyridines are extensively reported as unnatural metal ligands coordinating to metals such as Cu^{2+}, Zn^{2+}, and Fe^{2+}, ions of biological relevance.[4] More interesting are the changes in properties and function that occur when these peptides bind to metals. Inducing helical formation and/or nanoscale aggregation of metallo-peptides is described as how these subset entities differ in structure and function once assembled.

5.1.5 The Synthesis of Metal-Binding Peptides

Solid-phase peptide synthesis (SPPS) is a well-established chemistry, serving as a facile means for making these diverse set of compounds by a controlled step-wise incorporation of monomers. Natural amino acids building blocks are commercially available at relatively cheap costs. The relative inexpensiveness of this compound mitigates the inefficient consumption of the starting material to make peptides. However, when it comes to designing, synthesizing, and scaling up production of commercially unavailable, synthetically challenging building blocks, SPPS becomes a less attractive method for the introduction of unnatural residues. Amidation between the N-termini and/or Lysine residue with carboxylate-substituted bipyridines are used to overcome this challenge.

In addition, incorporation of a metal ligand does not guarantee that spontaneous folding will occur.[7] Therefore, devoting time and resources in the design of building blocks for solid-phase incorporation of unnatural residues may not be preferable when screening which functionalities cause peptides to fold. In fact, building block design of non-canonical side chains that can lead to peptide folding is regarded as an ongoing challenge in the field.[7] The alkylation approach previously described is seen as an alternative to building block

approaches when screening metal binding synthetic and folding of synthetic peptides. The commercial availability of alkyl bromide precursors as well as the synthetic ease to acquire brominated compounds makes alkylation strategies an attractive approach. From one sequence containing mono-methylated Lysine residues, different compounds can be incorporated, isolated, and studied for binding/folding properties. Once a side chain is found, building block design and synthesis for SPPS can be implemented as a means for expeditious synthesis of metal binding peptides, followed by optimization studies for improving folding, biological activity, etc.

The same precursor-based screening approach can be said of bipyridines with carboxylate functionalities. Lysine residues could be amide-coupled to different derivatives of bipyridine. However, isolation of the peptide may be more challenging due to the excess amount of coupling reagents required. Commonly reported solution-phase coupling conditions require the excess of EDC and NHS. Furthermore, selective modification is an additional complicating feature for solution-phase amidation. Unless the C-terminus is protected, concatenation will occur. In addition, the heterocyclic compounds like pyridine in the process may participate in unanticipated side reactions during the coupling.[10] In contrast, alkylation with bromide-containing precursors leaves impurities that can be readily removed with desalting or HPLC purifications.

5.1.6 Rationale for Alkylating Peptide with Bipyridine

Disordered peptides and proteins are induced to form folded structures upon coordination. Research points towards these common phenomena in nature to bring structurally disordered segments of protein, known as intrinsically disordered proteins (IDPs), into controlled folded patterns. This disorder-to-order transition key roles for diseases such as neurodegenerative disorders.[11] This served as inspiration for the work presented here. The hypothesis that the amine from the methylated Lysine would

preferentially bind to a metal as the third site of a tridentate ligand, further justifying the alkylation strategy. Particularly, bipyridine compound 3.2 would be best suited for this purpose (Figure 5.2). Compound 3.1 was the control to prove that positioning the methylated lysine on the ligand played a critical role. In other words, controlled folding of peptides would not readily occur, because the residue would not be preorganized to bind as a tridentate ligand.

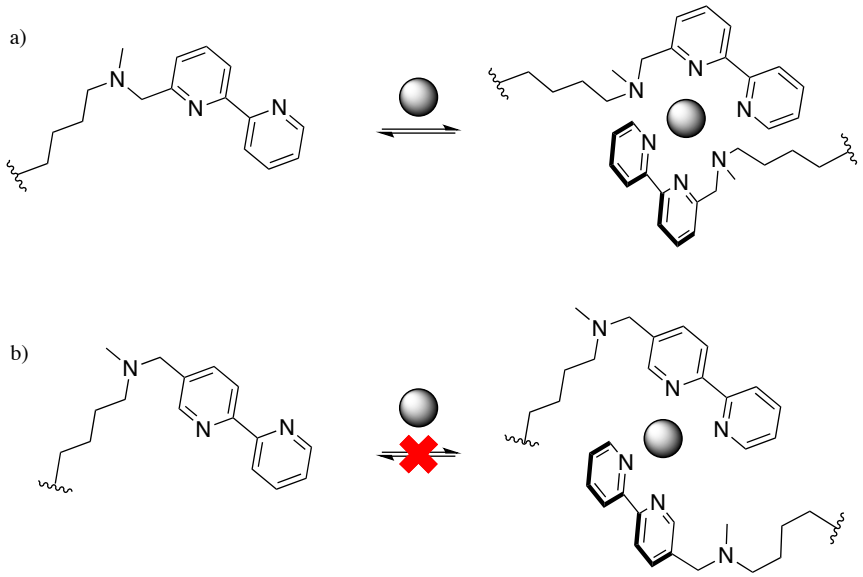
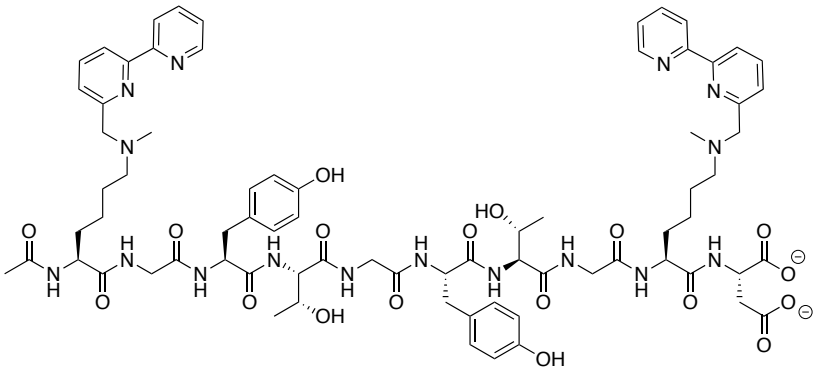
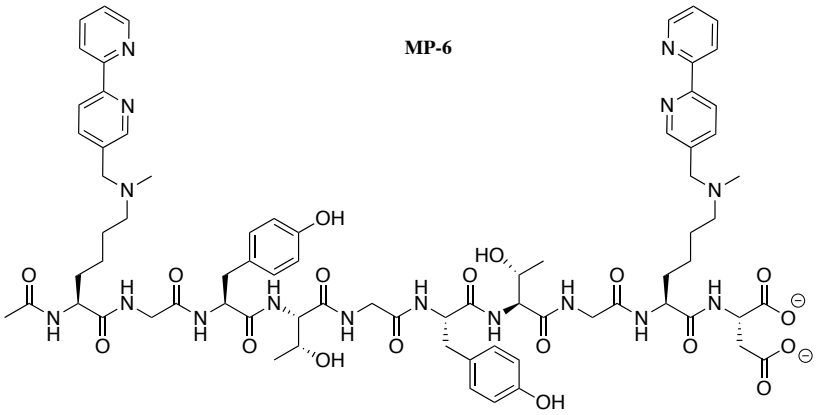


Figure 5.2: Proposed N-methyl Lysine modification with bipyridines. a) Compound 3.2 modified with methylated Lysine considered to bind and form a tridentate ligand binding around a metal center in a 2:1 stoichiometry. b) Compound 3.1 modified with methylated Lysine considered not to form a tridentate ligand.

In addition, bipyridines prefer coordination stoichiometries of 3:1 (ligand to metal center). Therefore, a peptide with only two ligands modified at position 6 (Figure 5.3) would favor cyclization. Peptide with ligands modified at position 5 (Figure 5.3) would not properly cyclize but participate in non-specific biding arrangements. A third ligand from an adjacent peptide would have to act as the third coordinating site, leading to a combination of different structures difficult to determine. These two peptides will be referred to as metallo-peptide (MP) 5 or 6, depending on the substitution of the bipyridine.



MP-6



MP-5

Figure 5.3: Metallo-peptides (MPs) used for cyclization studies.

5.1.7 Introduction of Dynamic Reversible Covalent Bonds

Once cyclization is confirmed, combining peptides with other unnatural functionalities that can participate in dynamic reversible covalent bonds is considered to mitigate issues that may arise when trying to induce folding of synthetic peptides with non-canonical functionalities. Recently, the Anslyn group reported a series of NMR studies confirming no cross-reactivity between four highly reversible covalent bonds: boronic acid esterification, thiol-conjugate addition, hydrazone formation, including supramolecular terpyridine complexation with a metal cation.[12] Bipyridine complexation is analogous to

the ligand studied previously, and the other three reactions can be incorporated as part of an ensemble to promote folding.

It is the dynamic reversibility of these covalent bonds formed that make the from these reactions attractive in the studies presented here. These kinds of reactions can be considered as an intermediate between the strong covalent bond, which usually locks precursors into a specific arrangement, and of non-covalent interactions, which tend to be too dynamic. If they are not biased by the principles of preorganization or complementarity, these kinds of supramolecular interactions tend to be form disordered entities within themselves or their immediate surroundings.[13] The work presented here focuses on peptides with aldehydes and hydrazides as the orthogonal functional groups combined with our bipyridine peptide cyclization studies.

So, these kinds of bonds are regarded to act as useful unnatural moieties when included in peptides. These will promote higher order rearrangement of simple peptide precursors. Because of their dynamic character, definitive targets are not envisioned. Instead, a set of simple components are mixed and the thermodynamically favored entity or entities are believed to form. In other words, like a wind-up toy, the mixing of components is the turning of the crank. The trajectory taken by the toy once it is released can be regarded as the products formed. The same can be said of proteins settling into their preferred conformational state.[14] In a broader sense, these biomolecules and organic synthetic structures can be referred to as "dynamers," because they are responsive to their environments and can change accordingly.[15] Interest for these kinds of bonds are a strong focus for accessing macroscopically responsive and dynamic polymers.[16] Such compounds can have potential pharmacological applications for introducing readily degradable compounds and minimizing toxic effects.[17]

5.1.8 Induce Cyclization of Bipyridine Peptides in the Presence of Aldehyde and Hydrazide Peptides and the Characterization of Higher-Ordered Structures

Therefore, this study's aim was to induce cyclization of MP-6 and MP-5 in the presence of an aldehyde peptide (AldPep) and hydrazide peptide (HydPep-1 and HydPep-2) shown in Figure 5.4. These peptides condensed upon [2,2'-bipyridine]-5,5'-dicarbaldehyde, and thus provided another ligand that could coordinate around a metal center. Although the formation of a thermodynamically favorable higher-order structure was not the goal, these studies were seen as preliminary ones that could serve as a basis for future work focused on the design and characterization of peptide libraries composed of these two orthogonal reactions.

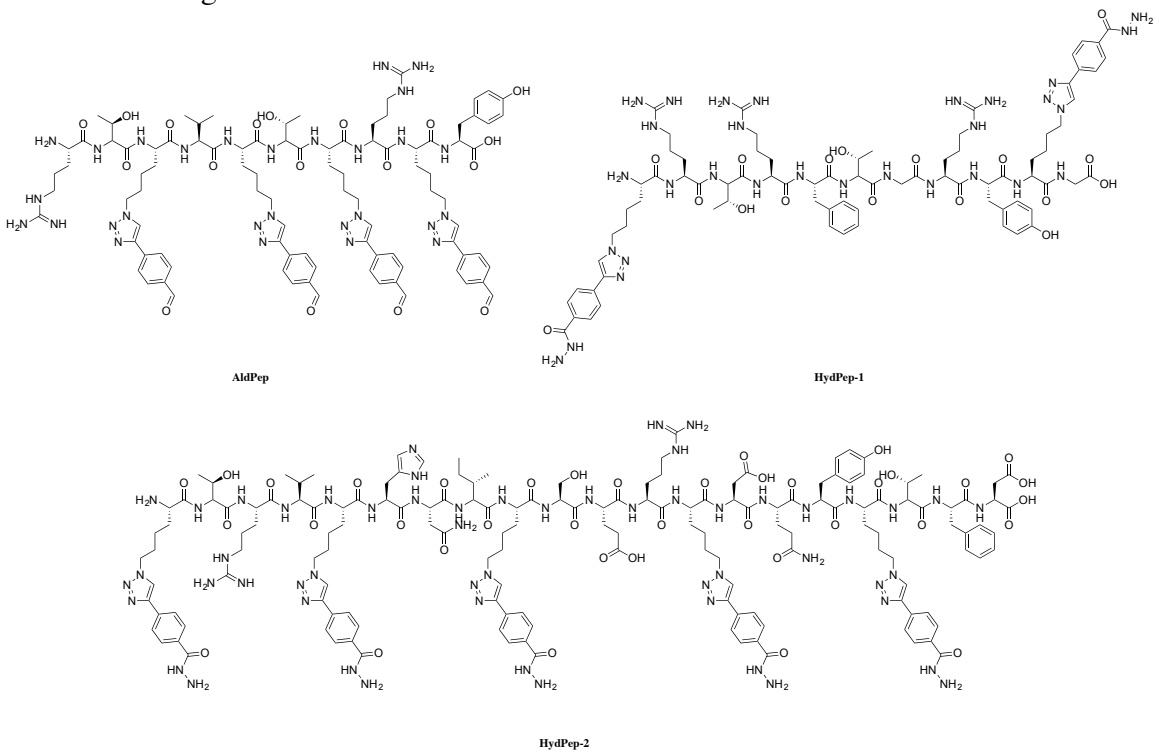


Figure 5.4: Hydrazone forming peptides used bipyridine peptide studies.

5.2 Results and Discussion

5.2.1 Rationale for MP-6 and MP-5

The sequence selected for MP-6 and MP-5 came from previously studied peptides from the Anslyn group that were used for cyclization. The peptides also contained a Tyrosine residue easy to identify at 280 nm during HPLC purification. The combination of Serine and Glycine increased hydrophilicity and minimized the steric repulsion from bulkier residues. Acetylated N-termini facilitated alkylation of secondary amine residues in solution. Acetylation also served as a precaution by removing basic sites that could compete to bind metals. As a result, residues such as Histidine and Arginine, known to bind to metal ions, were avoided.[18, 19] Length also played an important role. Peptides with short sequences were seen to bring residues in close contact with each other, disfavoring cyclization. Particularly, when non-covalent interactions were being used, such as metal complexation, bulkier residues on shorter peptides could impede desired folding.

5.2.2 Methods to Characterize Cyclization

Herein, we relied on UV-Vis and proton NMR titrations to characterize metal induced cyclization. After reaching saturation, the concentration of the titrant was most sensitive where complexation, was calculated. This value helped determine the correct stoichiometry for preferred folding. The Anslyn group previously has reported the use of these techniques for other assemblies.[20]

5.2.3 Titration Studies with Small Molecule Bipyridines

Initially, to validate our characterization methods, we performed titrations with 2-2’-bipyridine (5.3), 5-(bromomethyl)-2,2'-bipyridine (5.1), 6-methyl-2,2'-bipyridine (5.4), [2,2'-bipyridine]-5,5'-dicarbaldehyde (5.5), and [2,2'-bipyridine]-5,5'-diyldimethanethiol (5.6) (Figure 5.5). We tested substituted compounds to know what kind of functionalities were

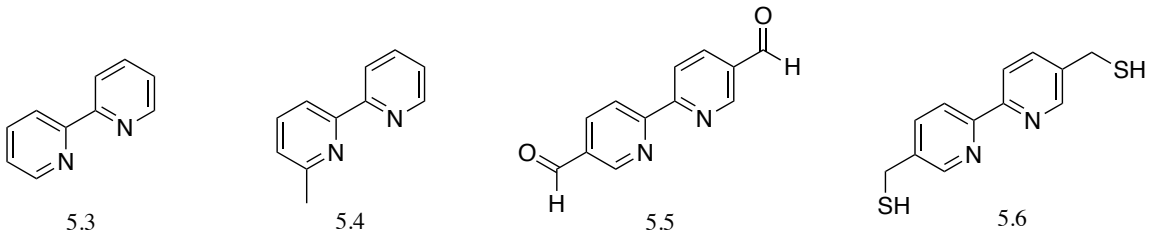


Figure 5.5: Small molecule bipyridines used for titration studies.

compatible in these studies. Compounds 5.1, 5.3, and 5.5 coordinated with Fe^{2+} in a 3:1 stoichiometry. Compound 5.6 also bound similarly with zinc, and compound 5.4, however, did not show any binding. The methyl group substituted at the 6-postion suggested that steric hindrance inhibited complexation. In addition, to observe any metal dependence for binding, Zn^{2+} was titrated into a solution of compound 5.4. Still, binding did not occur. At first glance, this result was potentially problematic. However, we believed possible substitution with peptides would allow greater flexibility for bipyridine arrangement around a metal. The hypothesis that incorporation of an amine from the methylated Lysine could act as the third member of a tridentate ligand would allowing for metal binding even when substituted at the 6-position.

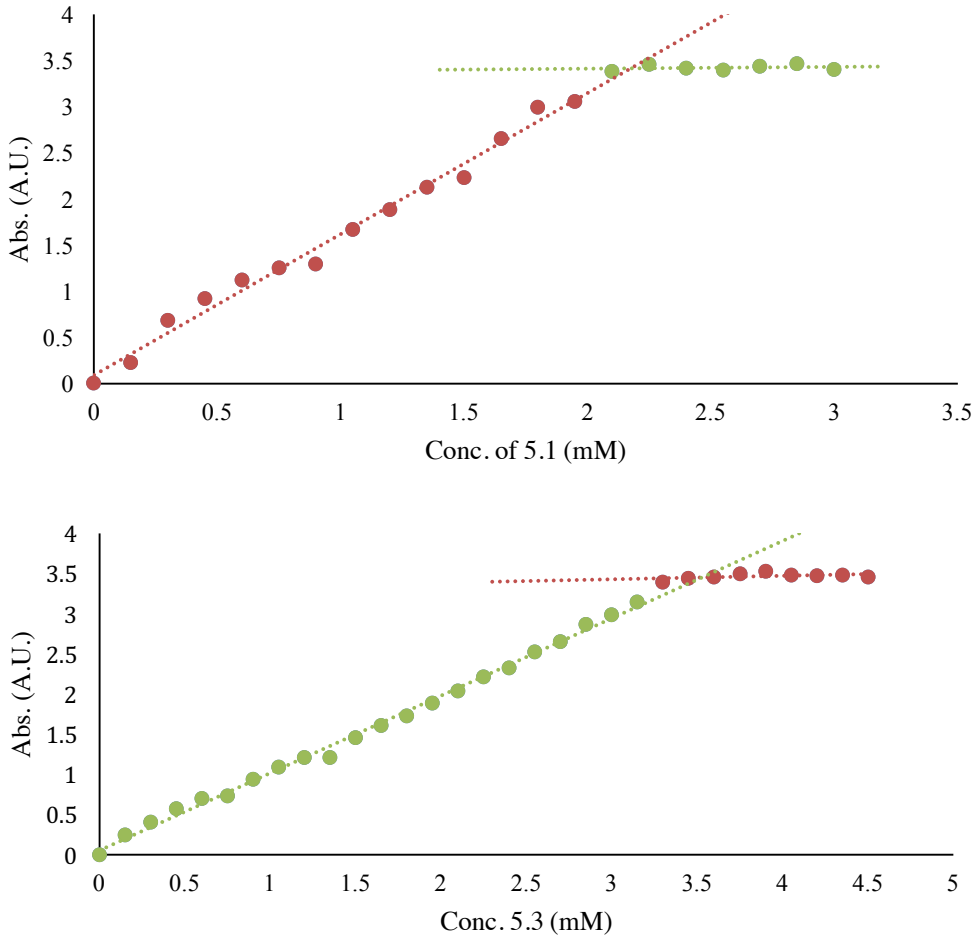


Figure 5.6: Binding isotherm for 5.1 (top) and 5.3 (bottom) at 530 nm.

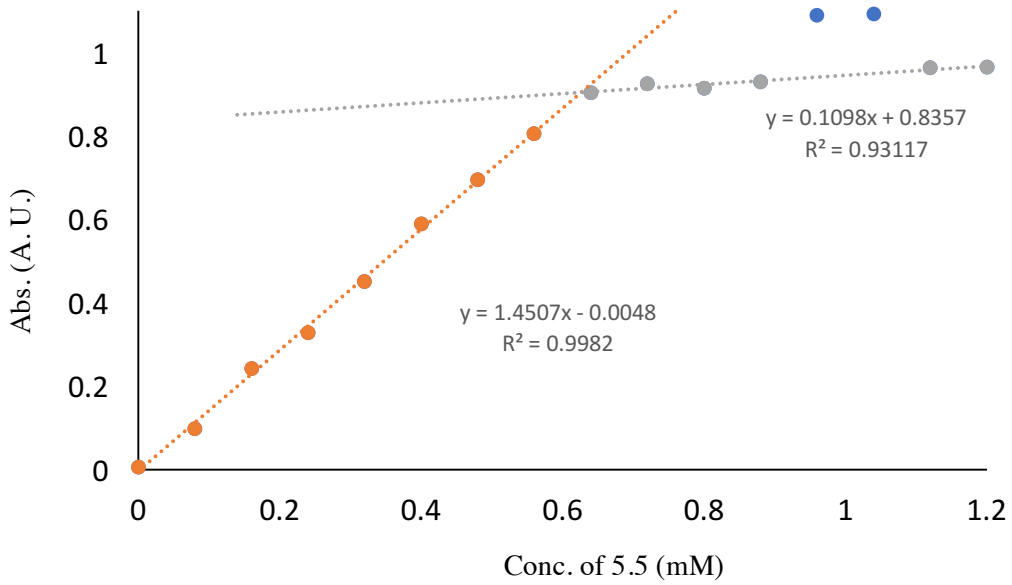


Figure 5.7: Binding isotherm for 5.5 at 530 nm.

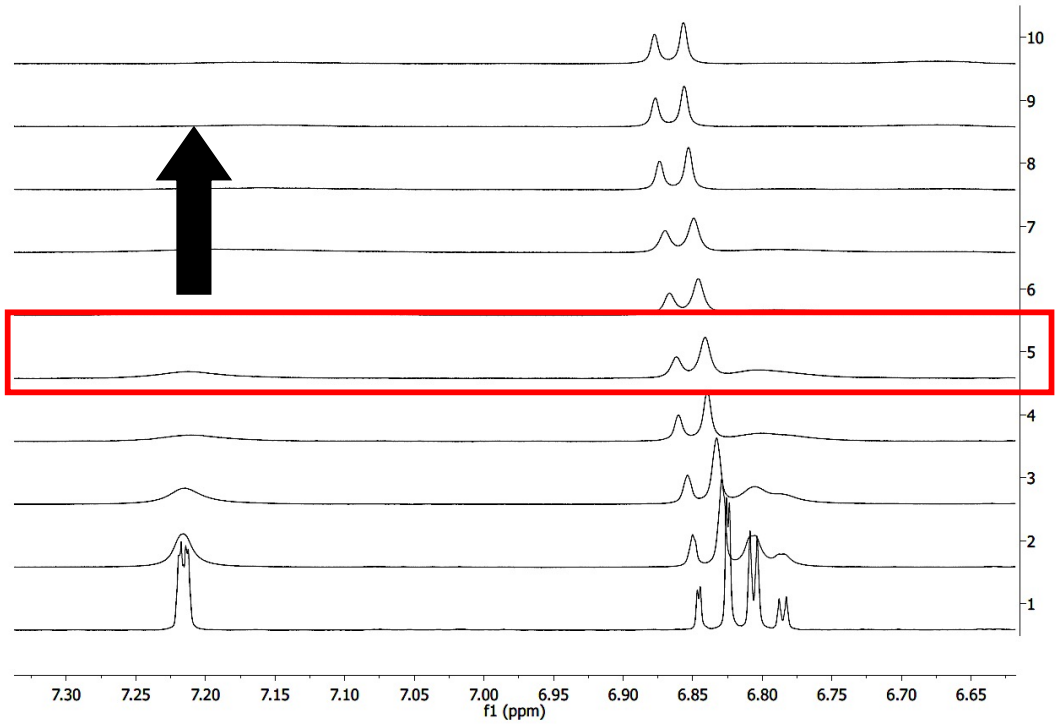


Figure 5.8: Proton NMR for Zn^{2+} titrated to a solution of 5.6. Red box indicates saturation reached at 0.33 eqs of Zn^{2+} added.

The binding of compound 5.1 demonstrated that bulkier substituents, like bromide, could be tolerated for ligands substituted at the 5-position. As a result, MP-5 was expected to bind readily.

5.2.4 MP-6 and MP-5 Binding with Fe^{2+} and Zn^{2+}

Titration studies for our peptides were initially characterized with proton NMR. To a solution of peptides, increasing amounts of metal were added. After each addition, successive NMR acquisitions tracked the alterations in the aromatic region, which is most sensitive to changes upon metal binding. Titrating Fe^{2+} to MP-6 led to a color change to a rusty red, but no notable changes were observed in the NMR spectra. Weakening of the signal intensity, however, was observed happened as the sample became more dilute (Figure 5.9). This suggested no metal was observed.

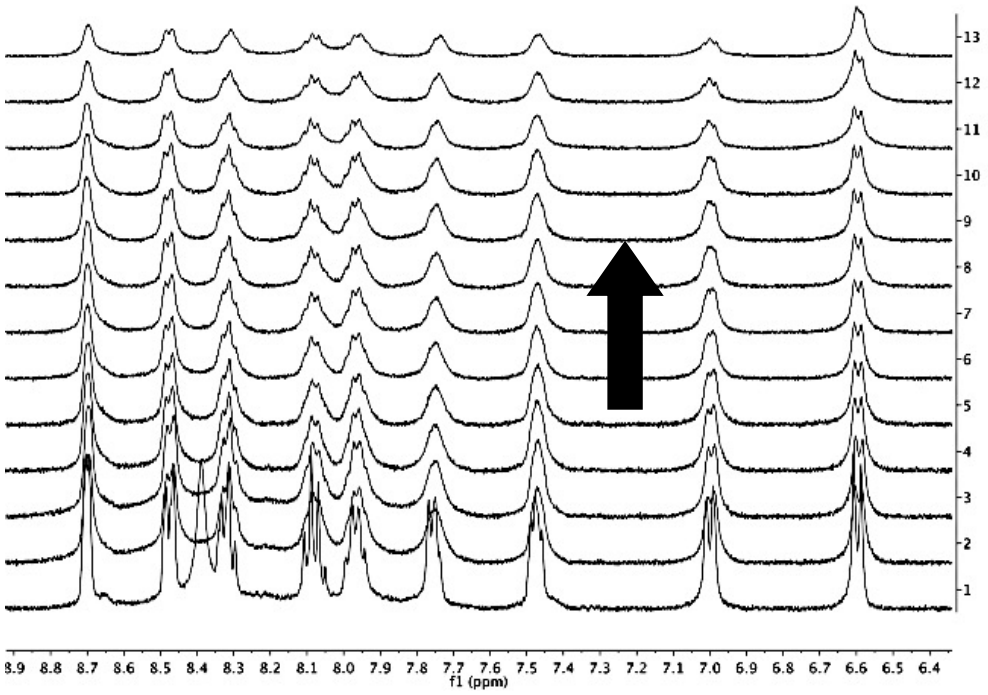


Figure 5.9: Proton NMR titration of Fe^{2+} to MP-6. Saturation was not reached.

In contrast, MP-5 was quite sensitive to the addition of Fe^{2+} (Figure 5.10). The signal, however, remained difficult to interpret in terms of the species formed. The NMR signals remained constant and were at their broadest at 1 equivalent of metal added. As the concentration of metal increased, additional peaks became prevalent, suggesting peptides were at their most oligomeric form. The more metal added, the more peptides began breaking apart into singly bound residues.

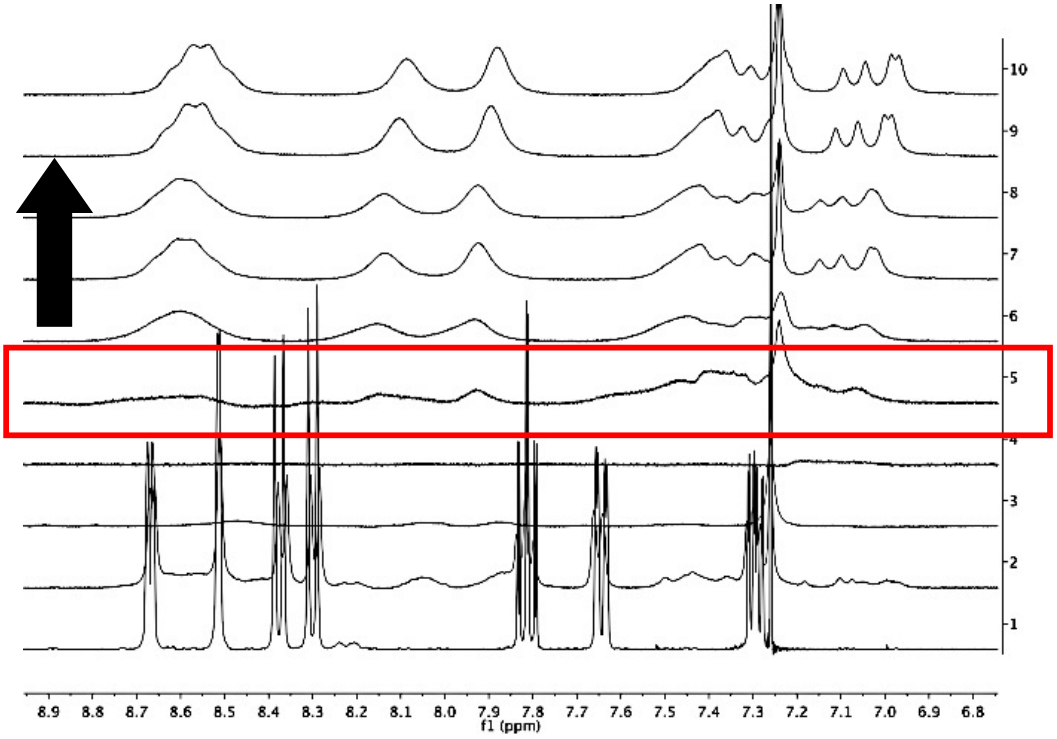


Figure 5.10: Proton NMR titration of Fe^{2+} to MP-5. Red box indicates 1 eq. of metal.

Zn^{2+} was also titrated and monitored by UV-Vis spectrophotometry into a solution of MP-5. No absorption change occurred, confirming steric hindrance at the sixth position maybe contributing to the lack of binding. Therefore, despite a potential coordination site introduced via N-methyl lysine, a 1:1 metal to peptide stoichiometry proved elusive. UV-Vis well-plate titration studies also demonstrated no response with iron (Figure 5.11).

Substitution at the 6-position colorimetric methylamine pendant increased the steric hindrance, impeding binding.

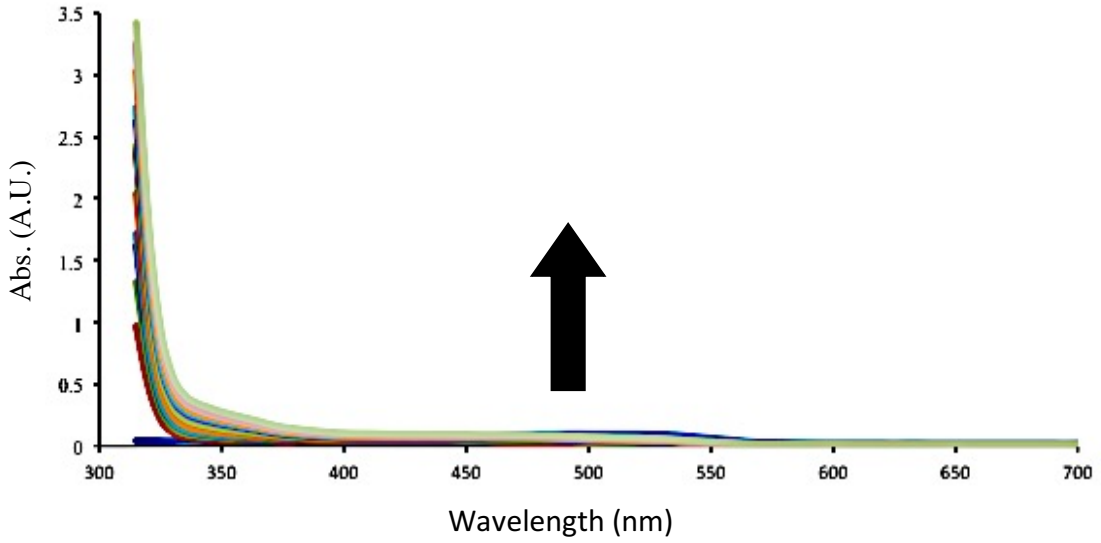


Figure 5.11: UV-Vis well-plate titration of MP-6 to Fe^{2+}. No colorimetric response was observed.

5.2.5 Ensuring Cyclization of MP-5

Since MP-6 did not bind as expected, further studies were not pursued. In contrast, MP-5 did respond to iron, as observed by UV-Vis spectrophotometry with an increase in absorbance. Literature accounts describing binding around metal centers with functionalized bipyridines and free 2-2'-bipyridine have been described, especially in the context of DNA binding studies with ruthenium centers.[21] So, well-plate titration studies with an equivalent amount of metal, peptide, and compound 5.3 took place. As shown in Figure 5.12, a well-behaved binding isotherm was acquired with this three-component ensemble, suggesting a more ordered and non-oligomeric folding in MP-5.. These titrations studies, repeated six times, confirmed a 1:1:1 stoichiometry. Similar to the small-molecule

titration studies. A total of three bipyridines were complexing around the cation. Two of the bipyridines came from the peptide and one from 2-2'-bipyridine in solution.

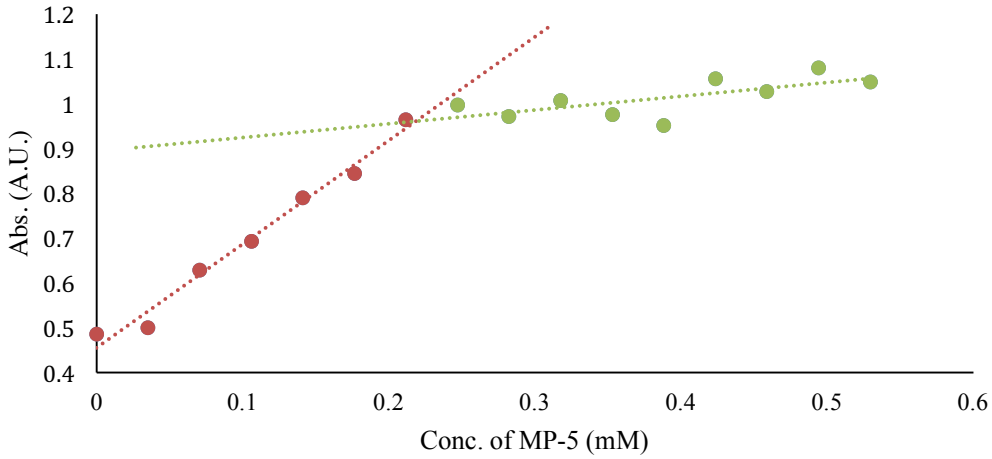


Figure 5.12: Binding isotherm for Fe^{2+} and compound 5.3 titrated with MP-5 at 530 nm.

5.2.6 Alternate Structure with Free Bipy: Oligomer

Because of the dilute concentrations, and the proximity of adjacent bipyridines on the peptide, cyclization was the most likely folding arrangement. Low concentrations tend to favor cyclic versus linear structures.[22] If oligomerization, or the association of multiple MP-5s, was occurring, proton NMR studies would have shown broadening of peaks in a similar manner to NMR studies done only with iron and MP-5. High-resolution mass spectrometry (HRMS) also corroborated cyclization. The fragmentation pattern revealed MP-5 bound to iron and in the presence of compound 5.3. Oligomeric usually indicated by m/z values differing by discrete values structures were not readily identified.[23]

5.2.7 MP Control Study

Peptide without bipyridine (MP) served as a control to rule out unintended binding. Aside from the binding occurring between the free bipyridine and, when increasing the concentration of MP, no change in absorption was observed, confirming that MP-5 indeed participated in complication.

5.2.8 Alternative Small-Molecules Bipyridines and MP-5

Since compound 5.3 successfully contributed to peptide folding, compounds 5.5 and 5.6 were also used for UV-Vis titration studies (Figure 5.13 & 5.14). From these studies, the desired 1:1:1 stoichiometry values were again confirmed, suggesting aldehydes and thiols do not interfere with complication.

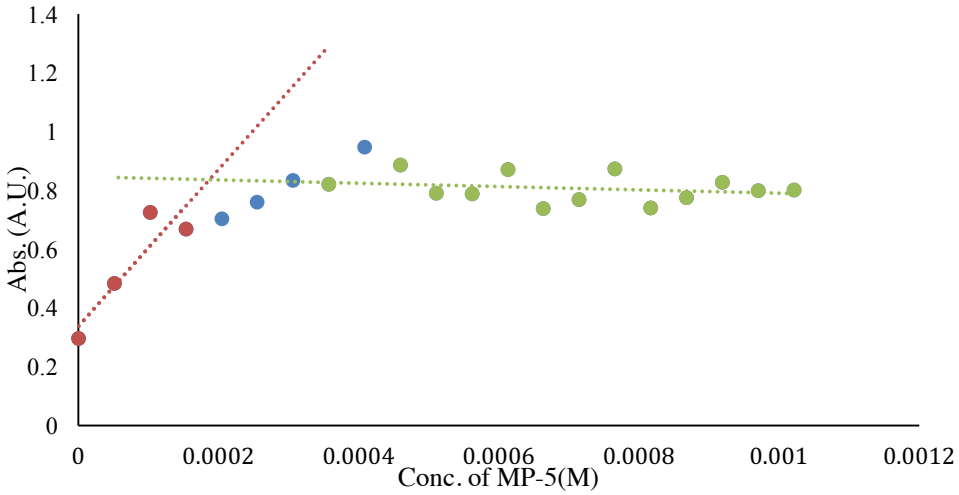


Figure 5.13: Binding isotherm for Fe^{2+} and compound 5.5 titrated with MP-5 at 530 nm.

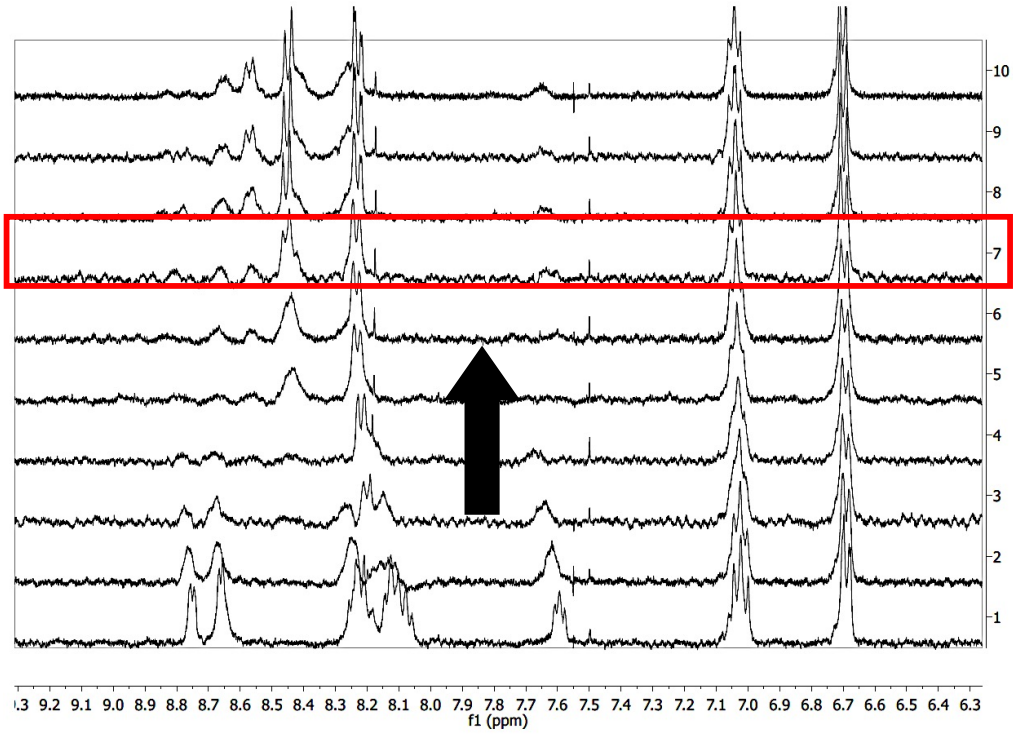


Figure 5.14: Proton NMR titration of Fe^{2+} to MP-5 and compound 5.6. Red box indicates1 eq. of metal.

5.2.9 Condensing HydPep-1 on Compound 5.5 and Complexation with MP-5

In our efforts to explore higher-order structure formation with orthogonal functionalities attached to small peptides, compound 5.5 served as an entry point. HydPep-1 readily complexed with the bipyridine-peptide derivative. The Anslyn group previously has studied this kind of peptide cyclization via hydrazone formation about a dialdehyde small molecule. These molecules readily form in aqueous mixtures at nuetral pH, making them ideal peptide partners with MP-5. The hydrazone cyclization formed readily in the presence of MP-5 with no observed cross-reactivity. Once cyclized HydPep-1 formed, the iron was titrated into the mixture, and the binding isotherm indicated 1:1:1 stoichiometry (Figure 5.16).

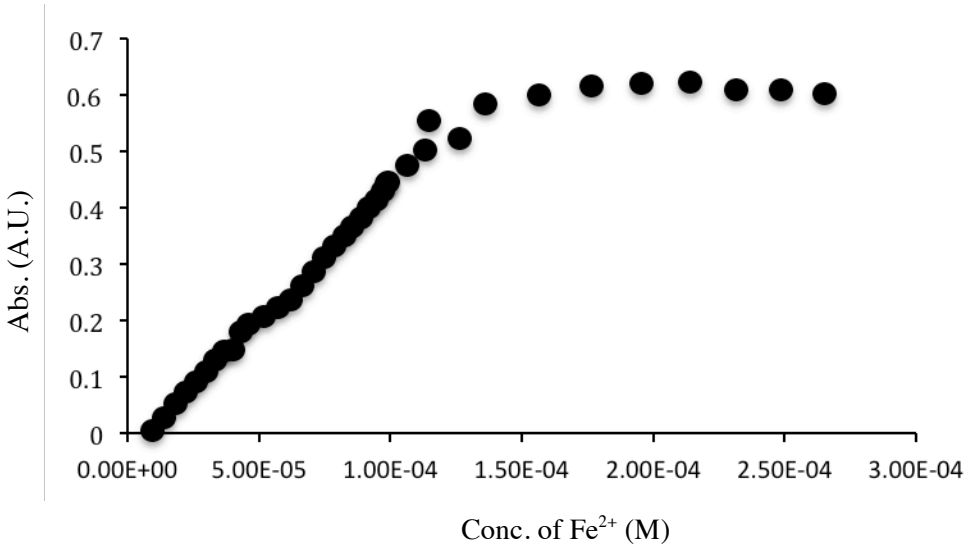


Figure 5.15: Binding isotherm for MP-5, cyclic HydPep-1, and Fe^{2+} at 545 nm.

Figure 5.17 is a representative structure proposed from these titration studies. Although the exact orientations of the peptides remained unconfirmed, mixtures of isomeric structures were believed to have formed.

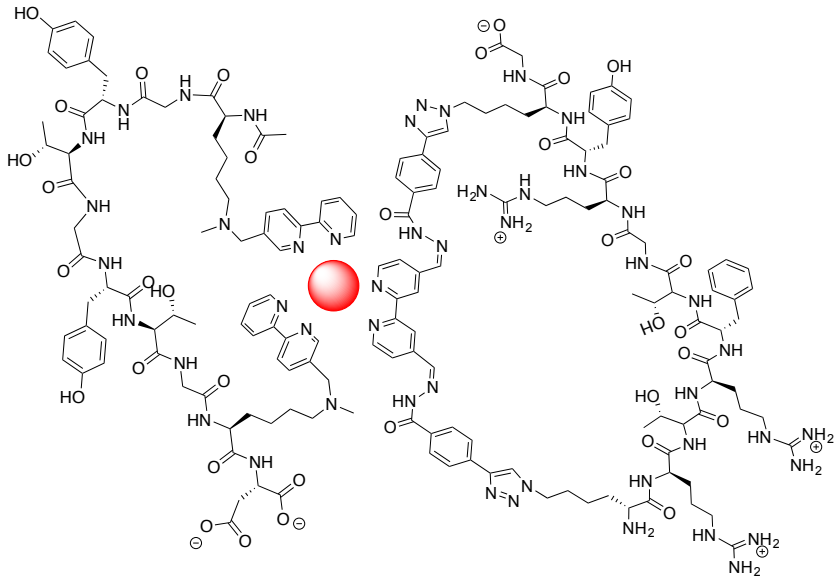


Figure 5.16: Representative structure formed between MP-5 and cyclic HydPep-1 around Fe^{2+}.

5.2.10 Condensing HydPep-2 on AldPep Followed by Condensing on Compound 5.5

To increase the level of complexity, HydPep-2 was condensed on AldPep. One hydrazide side chain remained unmodified upon reaction to further condense compound 5.5 onto the assembly. As with section 5.2.7, we did not determine the exact isomeric structure, however a mixture of products could have formed. These include structures with different C-termini orientations. Figure 5.18 is one possible structure that could form (compound 5.7). An attempt to condense two of these hydrazone peptide dimers did was not successful. In its place, a Tyrosine hydrazide served as an alternative. Further characterization is required to confirm this structure.

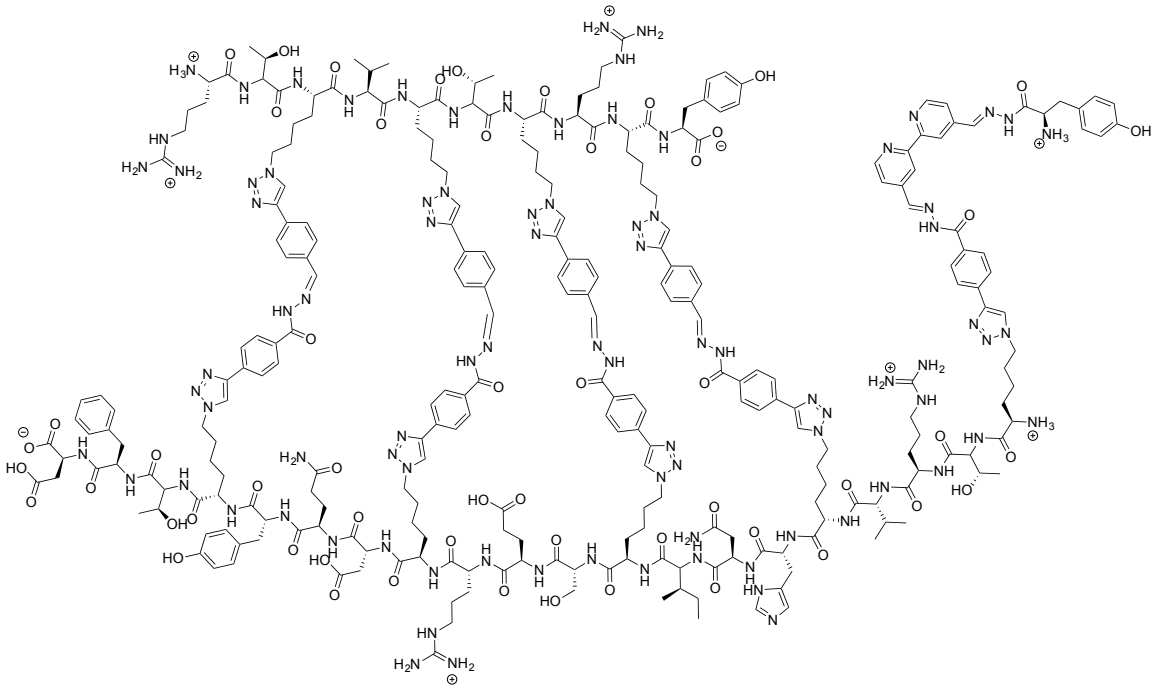


Figure 5.17: Representative structure formed with HydPep-2/AldPep dimer, hydrazide Tyrosine, and compound 5.5 (compound 5.7).

5.2.11 Complexation of Compound 5.7 with MP-5.

Although compound 5.7 was not confirmed, complexation appeared to occur with MP-5 and Fe^{2+} via UV-Vis titrations. Figure 5.19 proved saturation was reached. However,

it remained unclear if compound 5.7 formed part of the complex. A stoichiometry of 1:1:1 did prove promising, but a number of species could have been present. For example, only compound 5.5 may have participated. To relieve any strain, portions of AldPep could have dissociated to accommodate MP-5. Or, the hydrazide Tyrosine may have easily dissociated. Nevertheless, compound 5.8 was shown in Figure 5.20 to visualize the possibility of MP-5 being accommodated. MP-5 and the residue containing the hydrazone bipyridine oriented away from the bulky assembly, suggested the feasibility of this compound being formed.

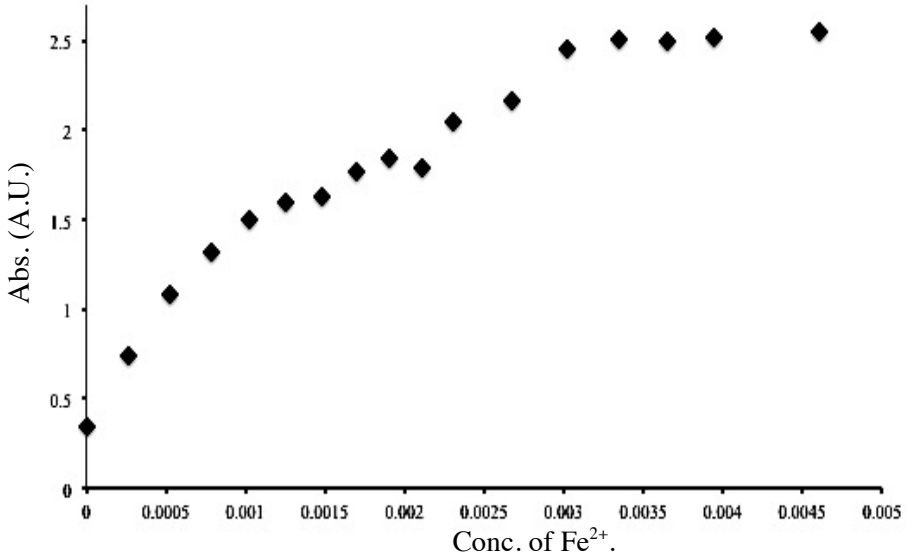


Figure 5.18: Binding isotherm for MP-5, Compound 5.7, and Fe^{2+} at 545 nm.

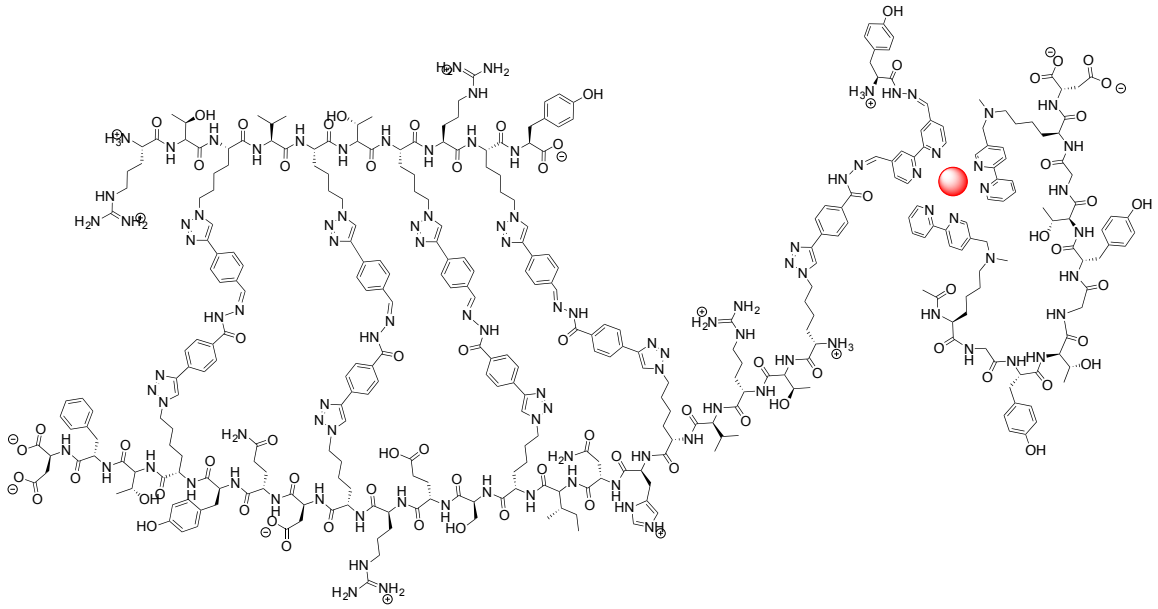


Figure 5.19: Possible structure formed when MP-5 and compound 5.7 bind around an Fe^{2+}.

5.3 Conclusions

Cyclization studies for MP-6 and MP-5 using proton NMR and UV-Vis titrations demonstrated that a free bipyridine was required for cyclic folding to occur. MP-5 readily folded, but MP-6 showed no affinity to iron or zinc. This suggested that steric hindrance at the 6-position from the methyl group impeded complexation. Compound 5.5 served as the necessary third bipyridine ligand that incorporated further complexity to higher order structures. HydPep-1 and HydPep-2 condensed to form novel bipyridine ligands that, when mixed with MP-5, suggested complexation in a 1:1:1 stoichiometry. Future studies will include determining molecular weights of these complexes via gel electrophoresis to confirm these large structures remain intact. Future studies will also include increasing the diversity of compounds formed by including other orthogonal reactions reported by the Anslyn group. Identifying the thermodynamically favored compound, or compounds, and their stereochemistry/regiochemistry will be the next challenge to solve in order to control folding of these synthetic structures.

5.4 Experimental

5.4.1 General Materials

For automated Fmoc amino solid-phase peptide synthesis, Gly, Thr(tBu), and Tyr(tBu) were purchased from P^3 Biosystems. Fmoc-Lys(N_ε,Me)-OH was purchased from Chem. Pep. Inc. Fmoc-Lys(N3)-OH was purchased from Chem-Impex, Inc. Fmoc-Gly-Wang resin (0.62 mmol g^{-1}) was purchased from P^3 Biosystems. N,N′-Diisopropylcarbodiimide (DIC) and Oxyma and (ethyl cyano(hydroxyimino)acetate) was purchased from Chem-Impex. DMF and piperidine used for automated solid-phase peptide synthesis were purchased from Fisher Scientific and Sigma-Aldrich. Ammonium acetate and acetic anhydride was purchased from Fisher. Methacrolein and 1-[2-oxo-2-(2-pyridinyl) ethyl] iodide, 6-Methyl-2,2′-dipyridyl, and [2,2'-bipyridine]-4,4'-dicarbaldehyde, and [2,2'-bipyridine]-5,5'-diyldimethanethiol[2,2'-bipyridine]-5,5'-diyldimethanethiol[2,2'-bipyridine]-5,5'-diyldimethanethiol were purchased from Sigma-Aldrich. N-bromo succinimide (NBS) and azobisisobutyronitrile (AIBN) were purchased from Acros Organics.

5.4.2 General Instrumentation

A Liberty Blue microwave peptide synthesizer was used for solid-phase peptide synthesis. Preparative HPLC purification of peptides was performed using an Agilent Zorbax SB-C18 Prep HT column 21.2 250 mm; 10 mL min1 , 5–95% MeOH (0.1% FA) in 90 min. Analytical HPLC characterization of peptides was performed using an Agilent Zorbax column 4.6 250 mm; 1 mL min1 , 5–95% MeCN (0.1% TFA) in 35 min (RT). A Gemini C18 3.5 micron 2.1 50 mm was used for online separation; 0.7 mL min1 , 5–95% MeCN (0.1% formic acid) in 12 min (RT). An Agilent Technologies 6530 Accurate Mass QTofLC/MS and a AB Voyager-DE PRO MALDI-TOF were used for high-resolution mass spectra of purified peptides. Solvents used were HPLC grade. A BioTek Cytation3

well plate reader and an Agilent Carey series spectrophotometer were used for UV-Vis titrations. A PowerEase Life technologies electrophoretic setup was used for gel separations of peptides and metallo-peptide complexes.

5.4.3 General Procedure for Peptide Synthesis of MP-5 and MP-6.

Synthesis of peptides was performed using standard settings using Liberty Blue software, and using 1 M of DIC and 1 M oxyma used as coupling and bases, respectively. Amino acid solutions were prepared at 200 mM, except for Fmoc-Lys(N_ε,Me)-OH which was prepared at 100 mM. Each peptide made was capped using acetic anhydride incorporated into the automated synthesis. After the synthesis, resin was washed with glacial AcOH (5 mL, 3x), DCM (5 mL, 3x), and MeOH (5 mL, 3x). Resins were placed under vacuum overnight. Peptides were cleaved from the resin using trifluoroacetic acid (TFA), triisopropylsilane, and nanopure water (95 : 2.5 : 2.5) (4 h). TFA was evaporated and the remaining oil was precipitated with diethyl ether at 0 °C. No further purification of the crude peptide was performed.

MP-6. Starting amount (0.016 mmole). Yield 38%. HRMS: $(M+2H)^{2+}$; calcd. 748.35390, obs. 748.35630.

MP-5. Starting amount: (0.015 mmole). Yield 25%. HRMS: $(M+2Na)^{2+}$; calcd. 770.33580, obs. 770.33800.

5.4.4 AldPep, HydPep-1, and HydPep-2

Dr. James Fredrick Ruether provided aldehyde and hydrazide peptides from the Anslyn group.

5.4.5 General Procedure for Alkylation of Peptides.

Peptides (1 eq.) were alkylated with 5-(bromomethyl)-2,2'-bipyridine (4.4 eqs.) or 6-(bromomethyl)-2,2'-bipyridine (4.4 eqs) in a MeCN, H_2O, and MeOH (80:15:5) (v/v/v) solvent mixture, followed by addition of 100 µL of Hunig's base (DIPEA). The reaction

was allowed to stir for 24 hrs at RT. The solution was purified by preparative HPLC. Purified samples were placed on the rotary evaporator to remove organic solvent. The aqueous remnants were frozen at -70 °C and lyophilized overnight.

5.4.6 Synthesis Brominated Ligand Precursors

Synthesis of 5-(bromomethyl)-2,2'-bipyridine was prepared using a protocol reported in the literature. [24]

Synthesis of 6-(bromomethyl)-2,2'-bipyridine was prepared using a protocol reported in the literature.[5]

5.4.7 Well-plate Titration Experiments[20]

Stock solutions of ligand peptides (0.6 mM), free bipyridine (0.2 mM), and Fe^{2+} (0.2 mM).

15 data points (15 wells) were taken for each experiment. To each well 25 μL metal and 25 μL of free bipyridine was introduced. Increasing amounts of peptide solution were introduced in 10 μL increments. The minimum was 0 μL solution introduced and the maximum was 150 μL introduced with all wells were diluted to a total of 150 μL. Each well was scanned between 300-700 nm. Blanks were composed of the MeOH:H_2O (1:1) (v/v).

Changes at 530 nm were monitored for titrations. A plot of concentration vs. change in absorption was made. Stoichiometry was determined by plotting a line along the increasing changes in absorption and a different line where the data points plateaued. The intersection was used to determine the amount of peptide required for optimal binding.

5.4.8 UV-Vis Titration for Hydrazone Peptides and MP-5 mixtures.

A titrant solution of Fe^{2+} (0.6 mM) was introduced prepared to a mixture of 0.2 mM hydrazone peptides and MP-5. This solution was titrated into a second solution of hydrazone peptide and MP-5 (0.2 mM) until saturation was reached.

5.4.9 NMR titration experiments.

NMR titrations were performed for both Zn^{2+} and Fe^{2+} to confirm UV-Vis titration experiments.

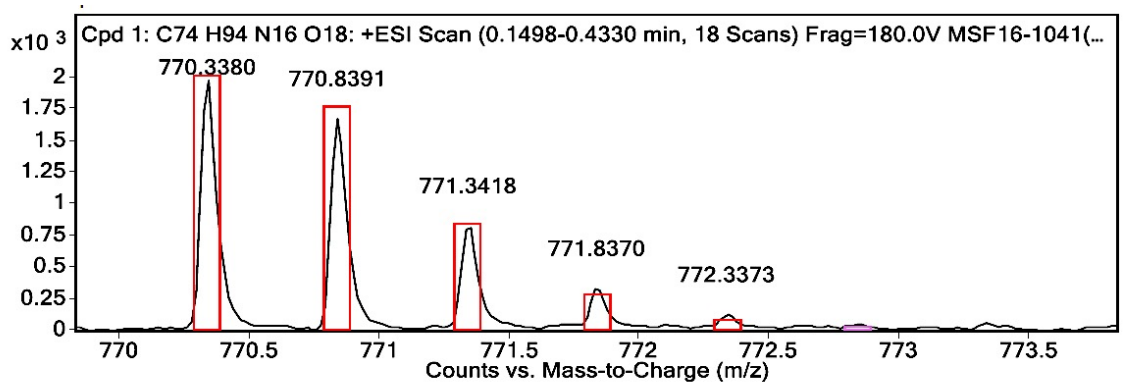
MP was dissolved to concentration in (2.3 μmole) 600 μL of. D_2O:MeOD (1:1) (v/v). Zinc or iron triflate was dissolved 180 μL (4.6 μmole) of deuterated solvent. 600 μL of the peptide solution was introduced to the NMR tube. An NMR was taken initially before the incorporation of any metal. Increment additions were added of metal solution (15 μL, 6x; 30 μL, 3x). After each addition, an NMR experiment was performed, immediately after. Titration of unreacted bipyridines was performed in a similar fashion.

5.5 Acknowledgements

We thank our following funding sources: the National Institute of Health (5DP1GM106408), the Defense Advanced Research Projects Agency (DARPA, N66001-14-2-4051), and the Welch Regents Chair (F-0046).

We also thank Anslyn undergraduate researchers Sang-Yop Kwon and Jonathan Partridge for synthesizing MPs and for the hours spent doing NMR titrations.

5.6 Experimental Characterization Data



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
499.24200			22563.5			
770.33800	770.33580	2	1978.04	C74H94N16O18	(M+2Na)+2	-2.81
770.83910	770.83730	2	1684.36	C74H94N16O18	(M+2Na)+2	-2.29
771.34180	771.33870	2	837.47	C74H94N16O18	(M+2Na)+2	-3.95
771.83700	771.84010	2	342.13	C74H94N16O18	(M+2Na)+2	4
772.33730	772.34140	2	134.56	C74H94N16O18	(M+2Na)+2	5.31

Figure 5.20: HRMS for MP-5.

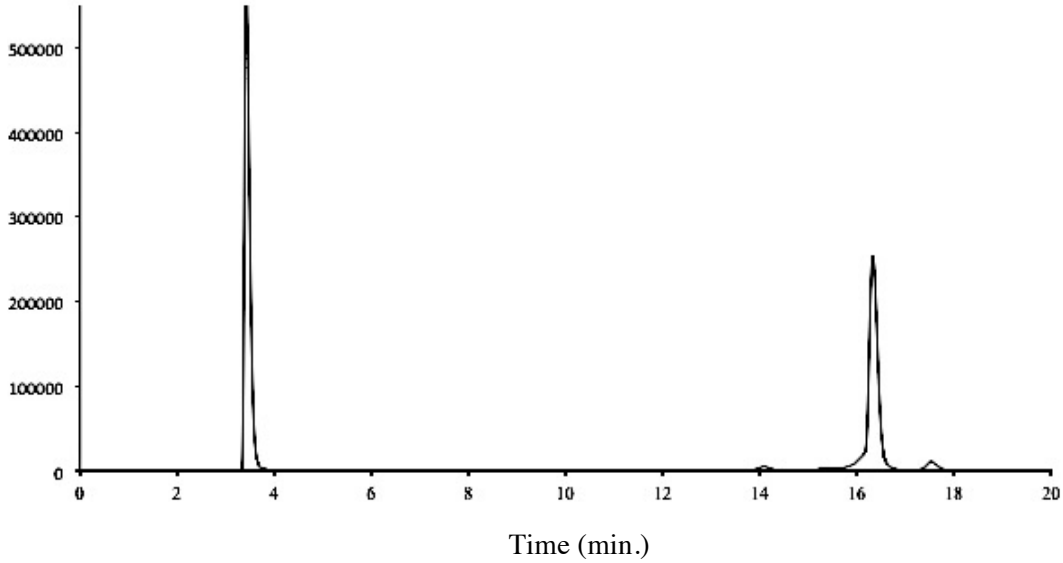
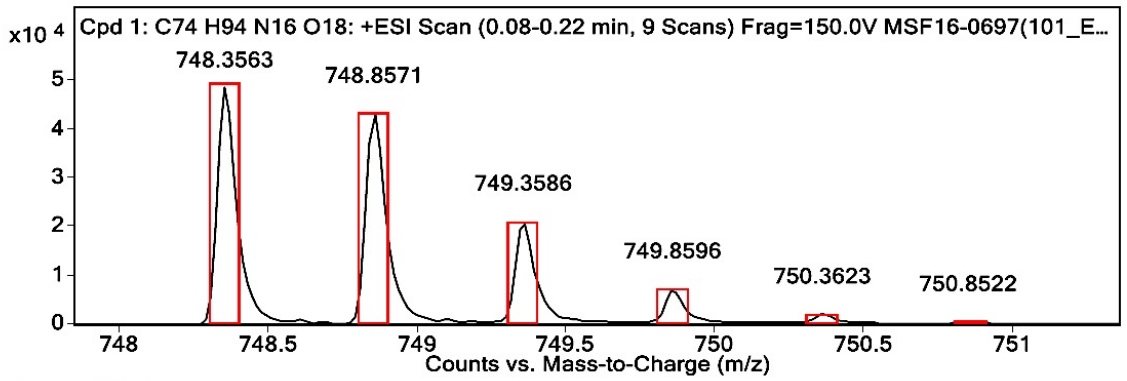


Figure 5.21: Analytical HPLC trace for MP-5 at 254 nm.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
499.24110			153640.8			
748.35630	748.35390	2	48860.6	C74H94N16O18	(M+2H)+2	-3.26
748.85710	748.85540	2	43099.37	C74H94N16O18	(M+2H)+2	-2.37
749.35860	749.35680	2	20948.22	C74H94N16O18	(M+2H)+2	-2.44
749.85960	749.85810	2	7048.33	C74H94N16O18	(M+2H)+2	-1.99
750.36230	750.35950	2	1867.83	C74H94N16O18	(M+2H)+2	-3.76
750.85220	750.86080	2	316.37	C74H94N16O18	(M+2H)+2	11.44

Figure 5.22: HRMS for MP-6.

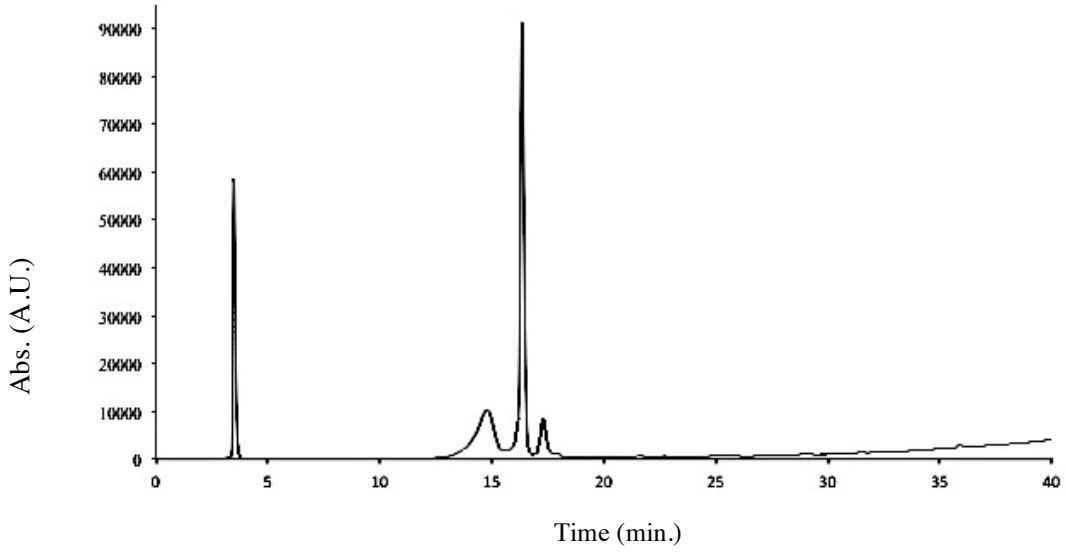
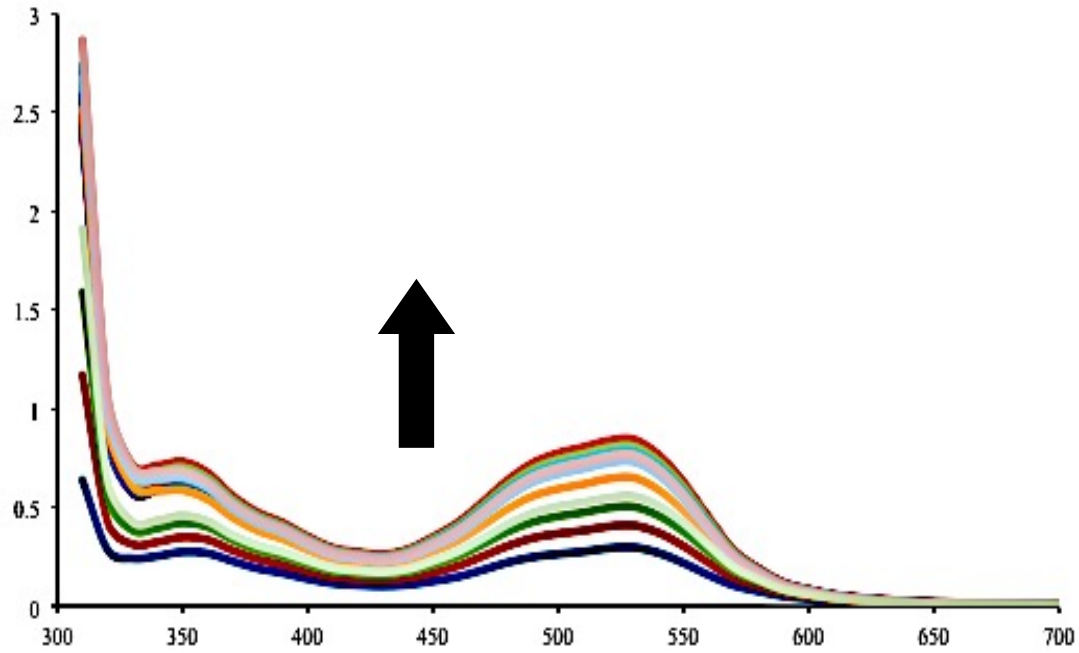


Figure 5.23: Analytical HPLC trace for MP-6 at 254 nm.



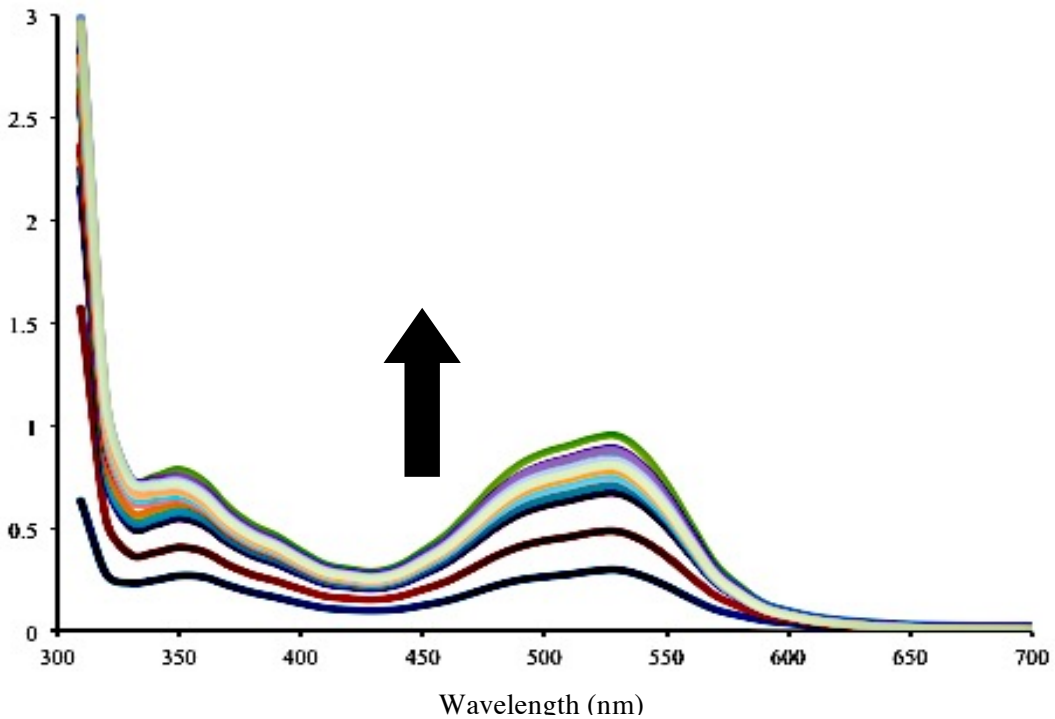


Figure 5.24: (top) Well-plate titration for 2-2' bipyridine and Fe^{2+} (trial 1). (bottom) Trial 2.

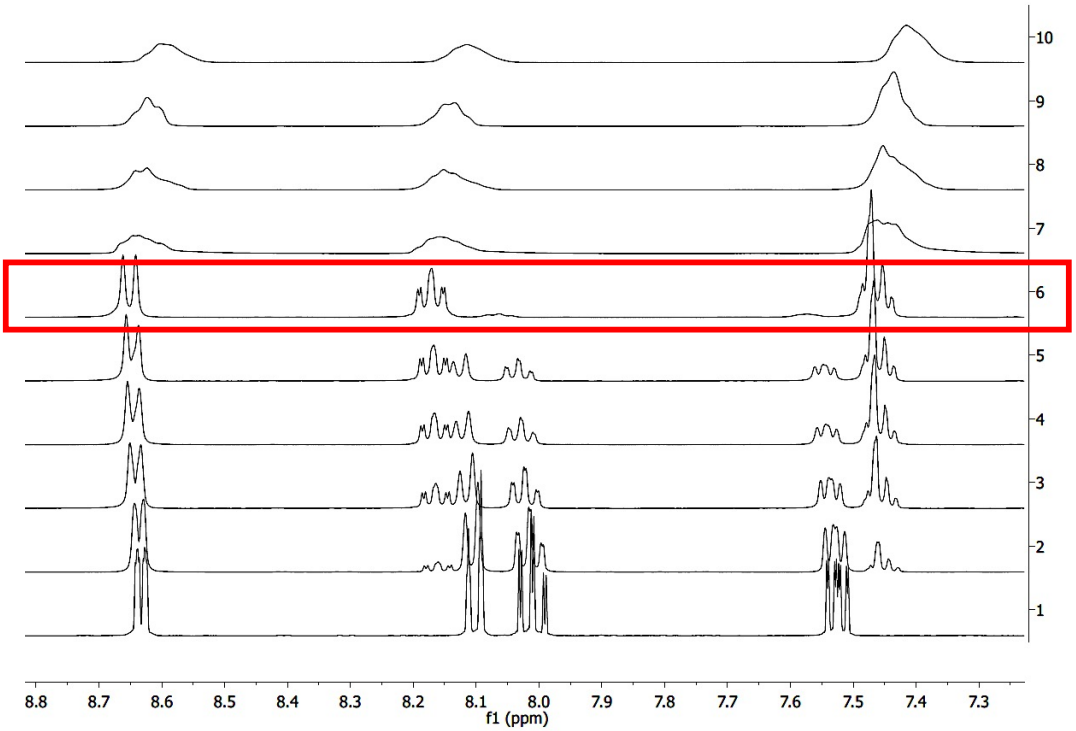


Figure 5.25: Proton NMR titration of 2-2' bipyridine with Fet^{2+}. Red box indicated 0.32 eq. required of metal required for 3:1 stoichiometry.

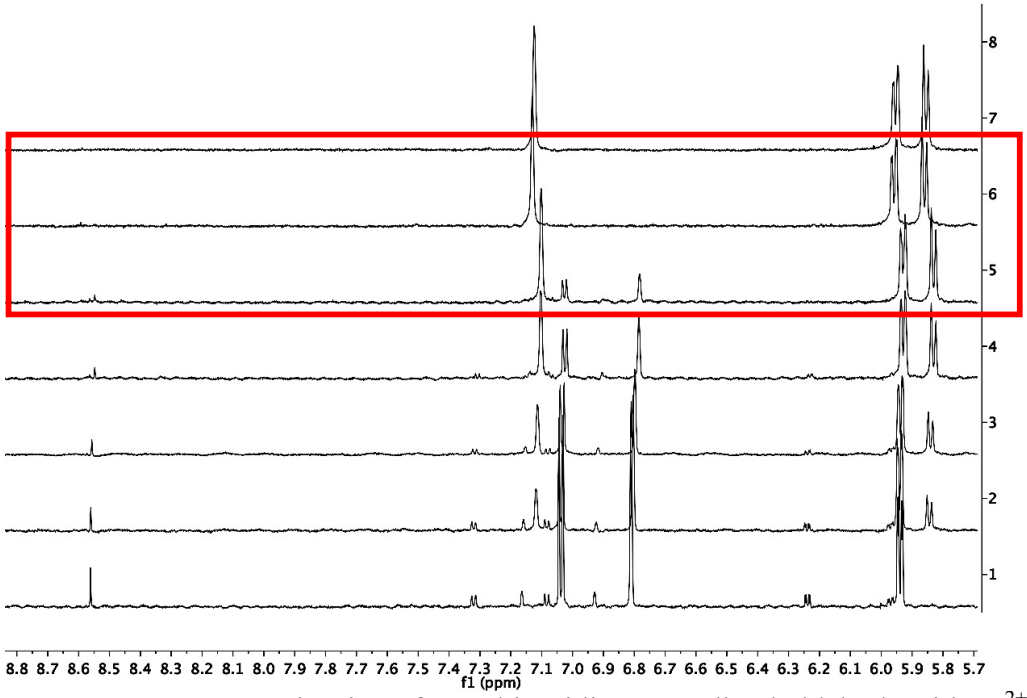


Figure 5.26: Proton NMR titration of [2,2'-bipyridine]-5,5'-dicarbaldehyde with Fe^{2+}. Red box indicated 0.32-0.42 eq. required of metal for 3:1 stoichiometry.

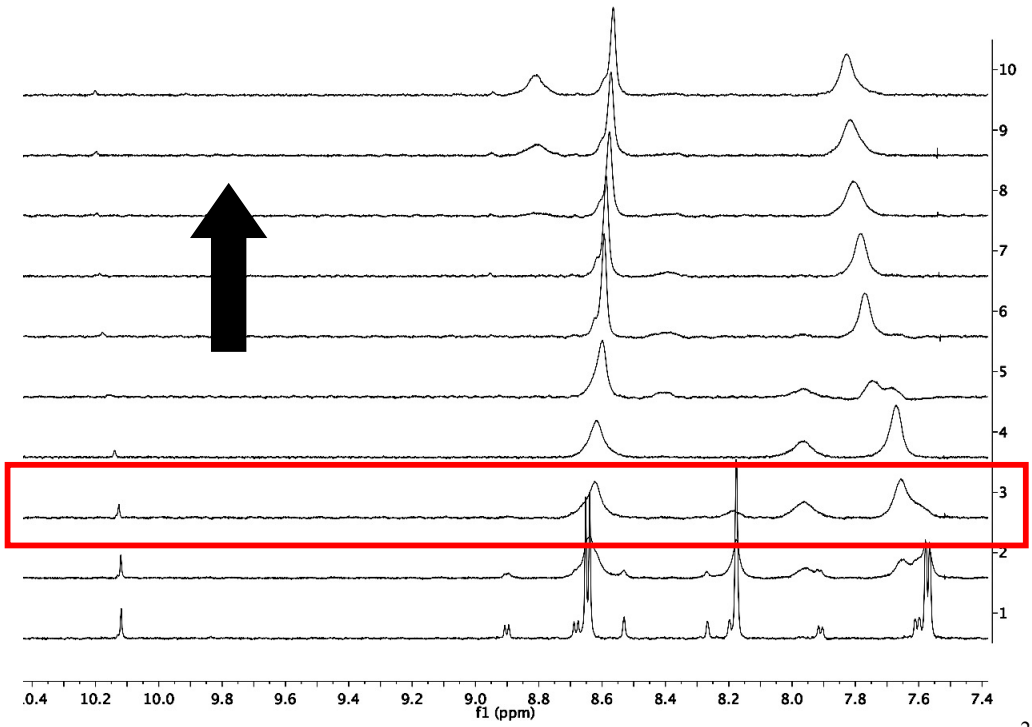


Figure 5.27: Proton NMR titration of [2,2'-bipyridine]-5,5'-dicarbaldehyde with Zn^{2+}. Red box indicated 0.33 eq. required of metal for 3:1 stoichiometry.

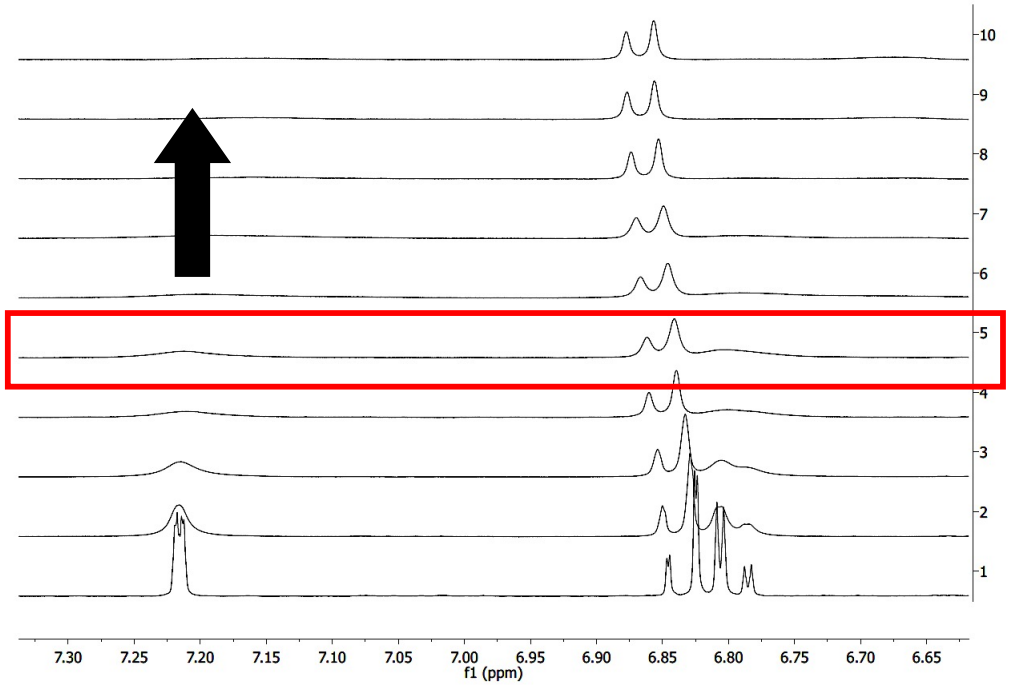


Figure 5.28: Proton NMR titration of [2,2'-bipyridine]-5,5'-diyldimethanethiol with Zn^{2+}. Red box indicated 0.33 eq. required of metal for 3:1 stoichiometry.

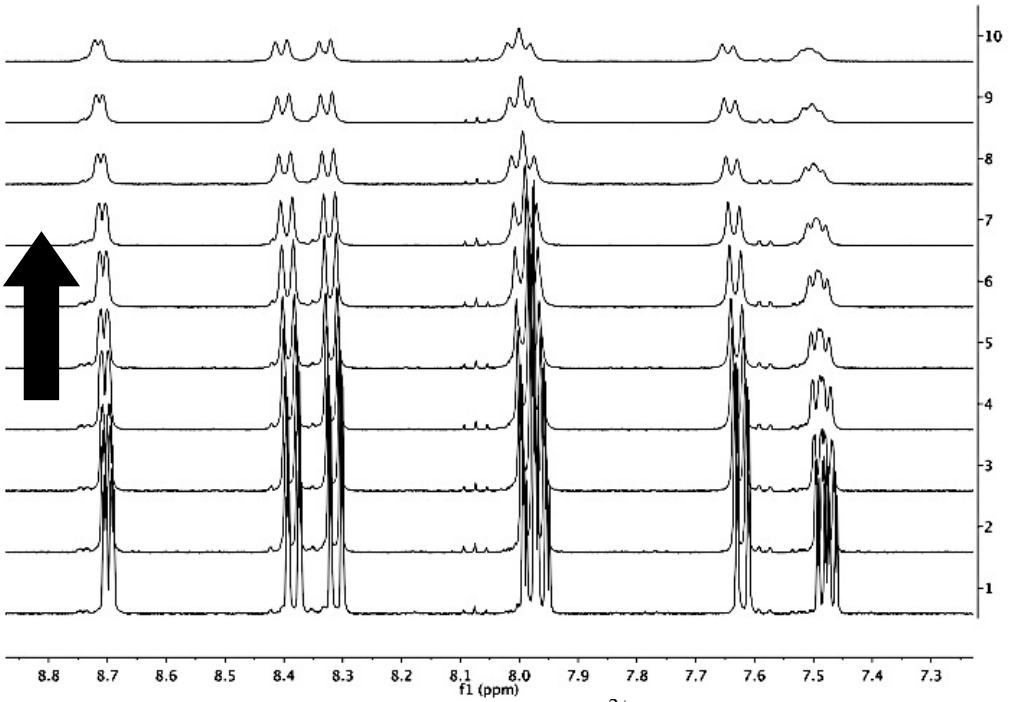
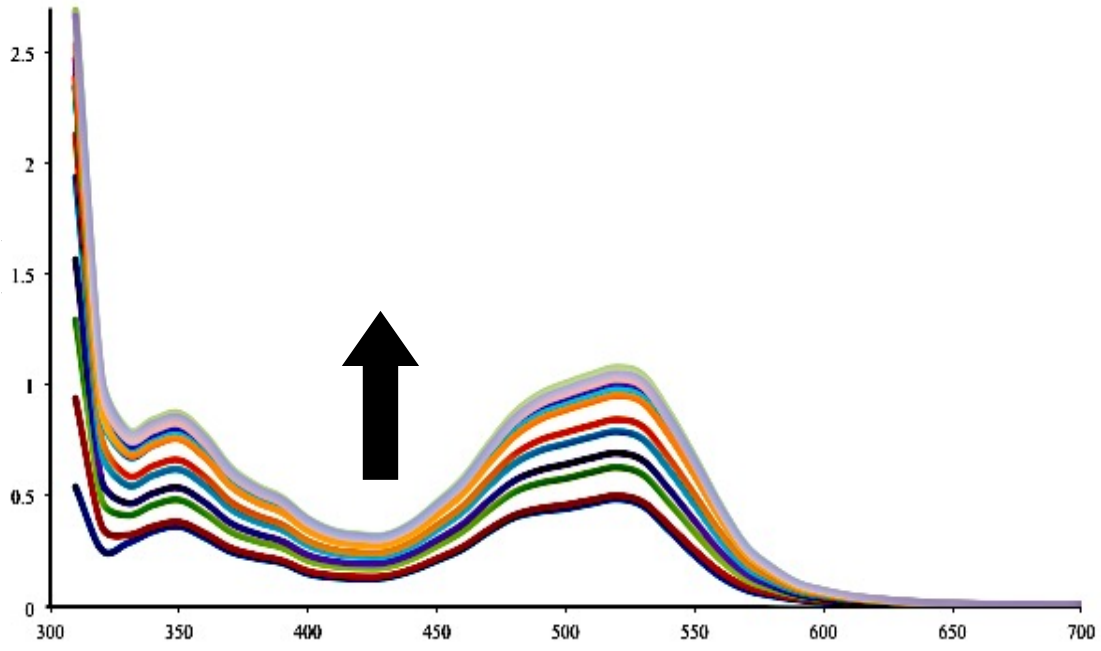


Figure 5.29: Proton NMR titration of MP-6 with Fe^{2+}. Saturation was never reached.



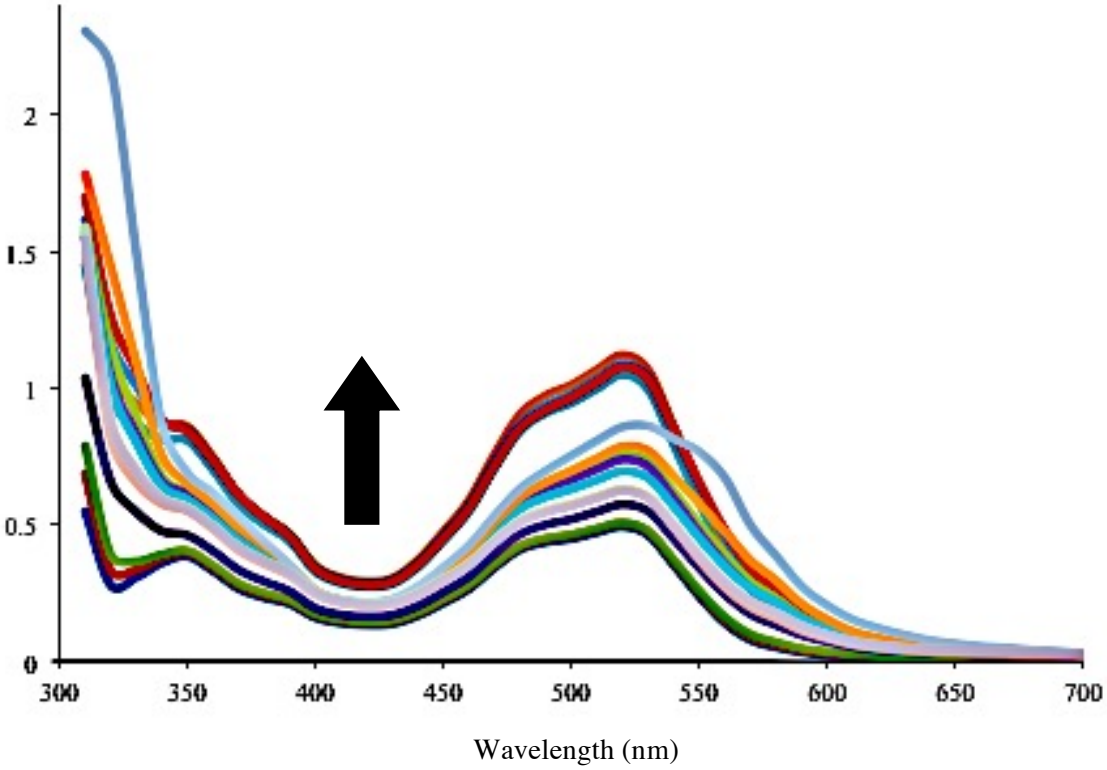
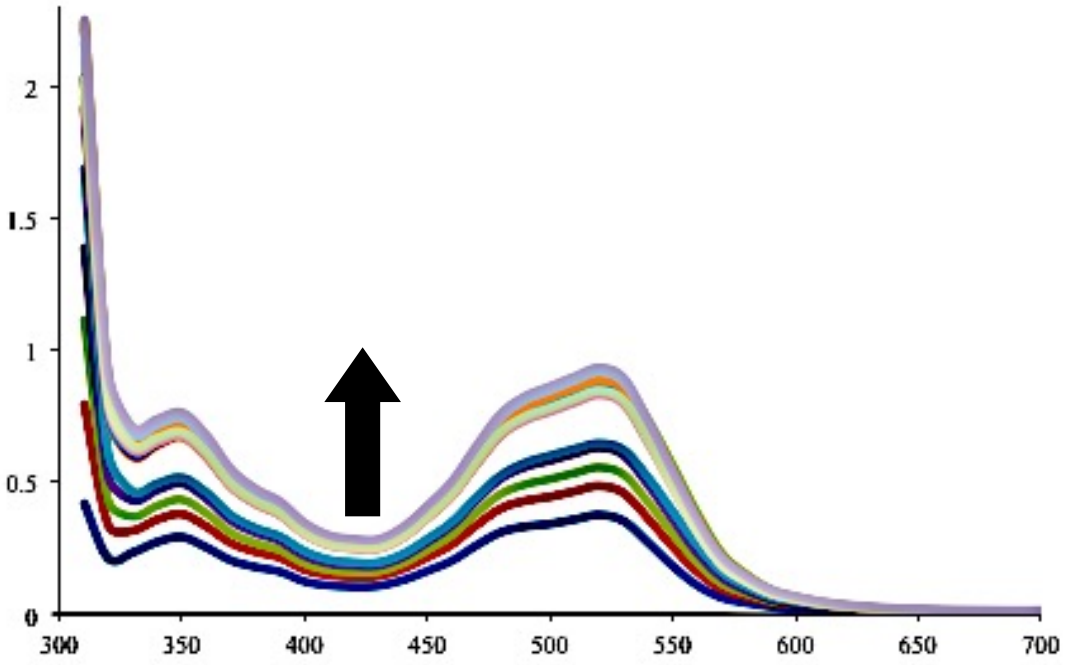


Figure 5.30: (top) Well-plate titration for MP-5, 2,2'-bipyridine and Fe^{2+} (trial 1). (bottom) Trial 2.



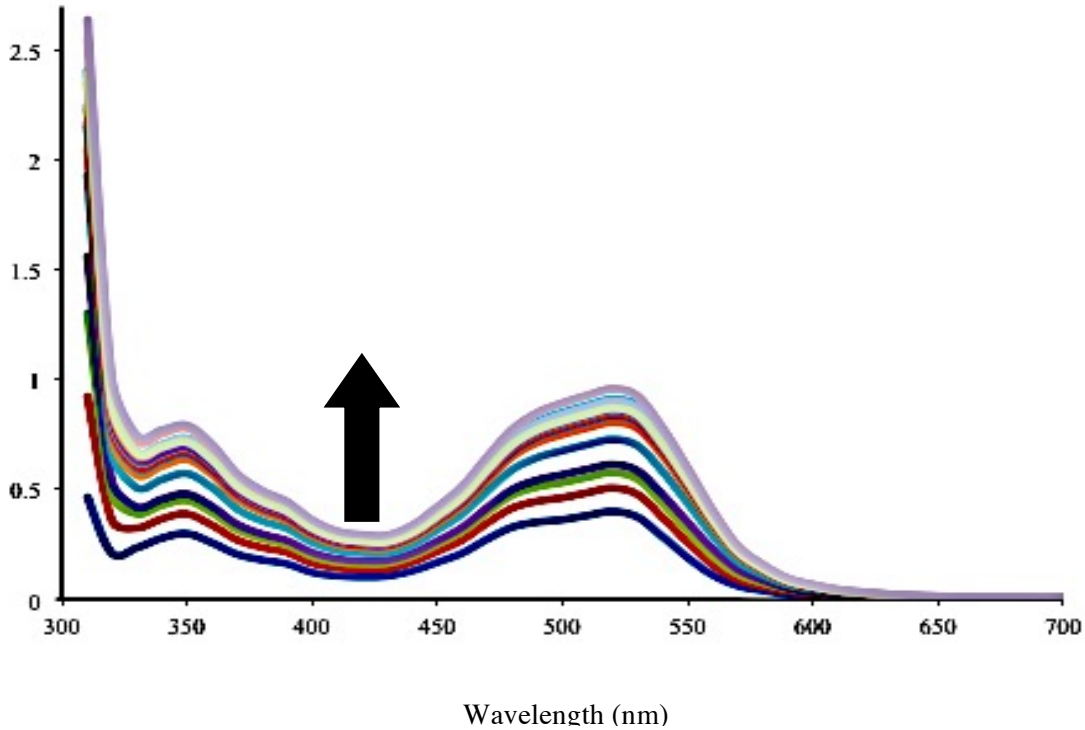
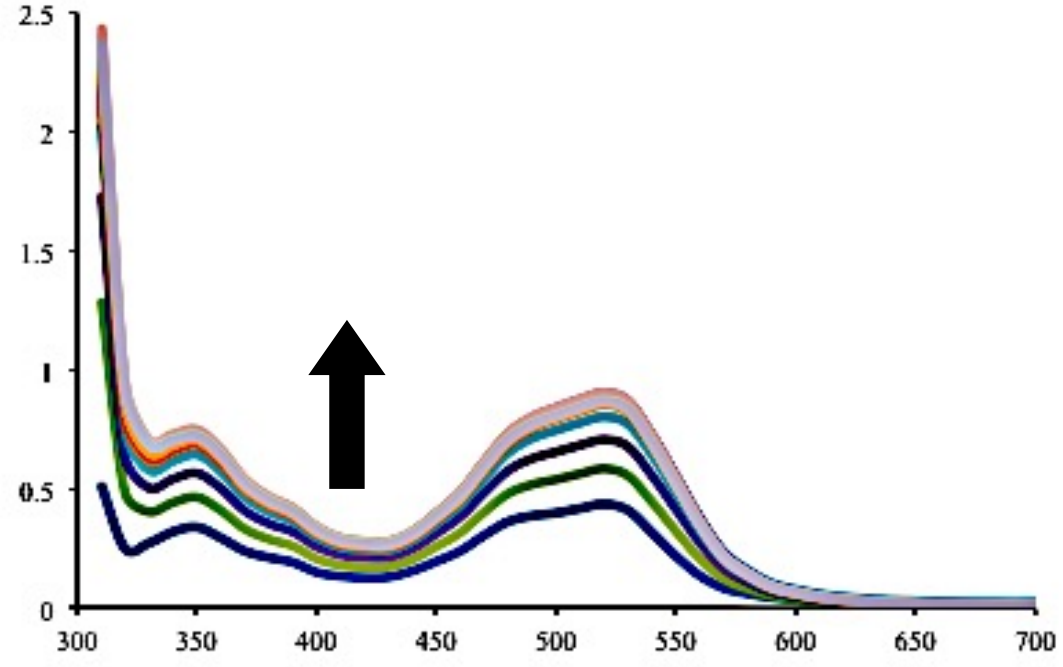


Figure 5.31: (top) Well-plate titration for MP-5, 2,2'-bipyridine, and Fe^{2+} (trial 3). (bottom) Trial 4.



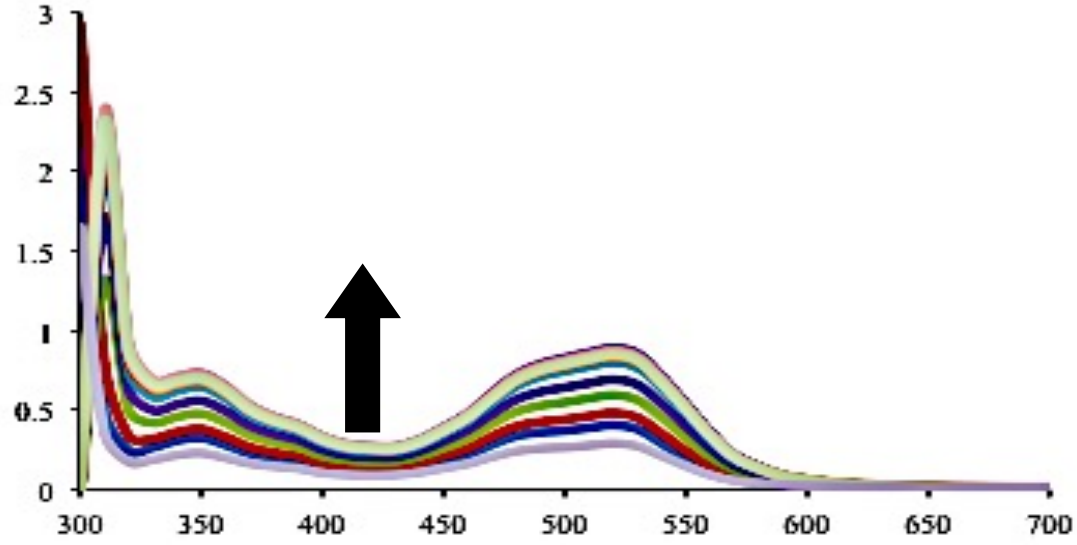


Figure 5.32: (top) Well-plate titration for MP-5, 2,2'-bipyridine, and Fe^{2+} (trial 5). (bottom) Trial 6.

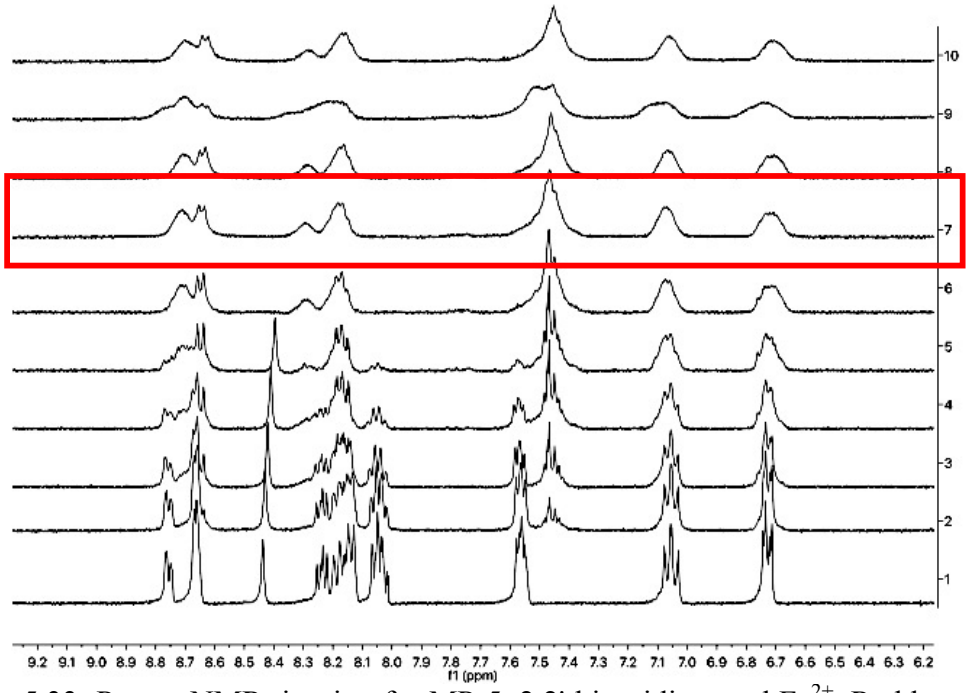


Figure 5.33: Proton NMR titration for MP-5, 2,2'-bipyridine, and Fe^{2+}. Red box indicates 1.25 eqs of metal were required for cyclization.

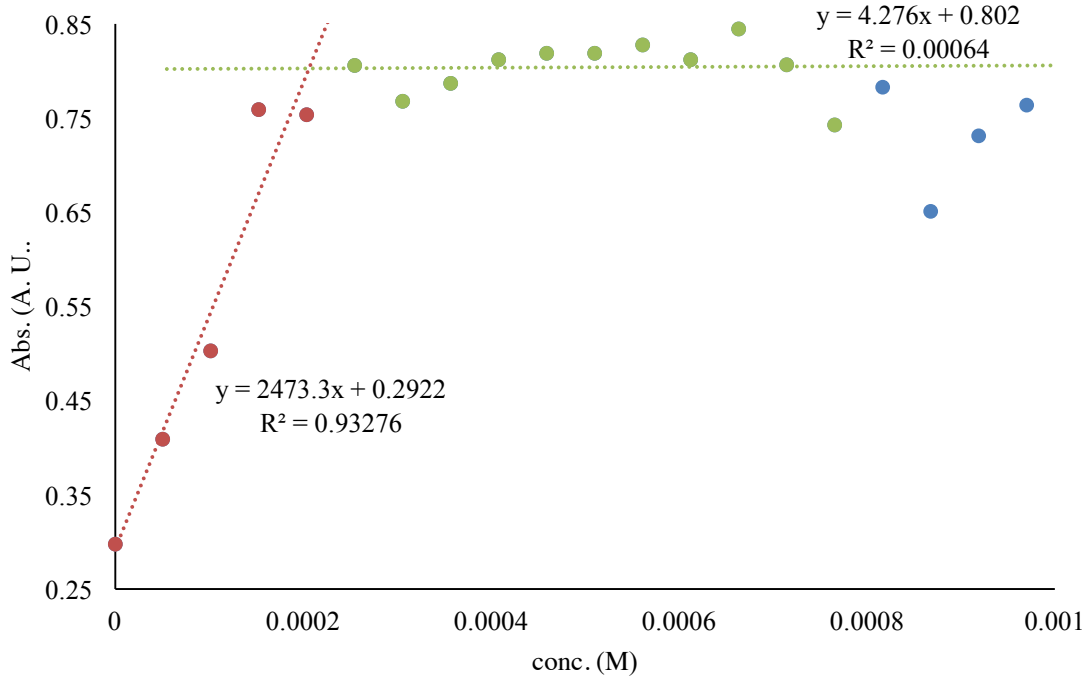


Figure 5.34: Binding plot at 530 nm for MP-5 (0.206 mM), [2,2'-bipyridine]-5,5'-dicarbaldehyde (0.195 mM), and Fe^{2+} (0,214 mM). Cyclization occurred with a 1.1:1:1.1 stoichiometry.

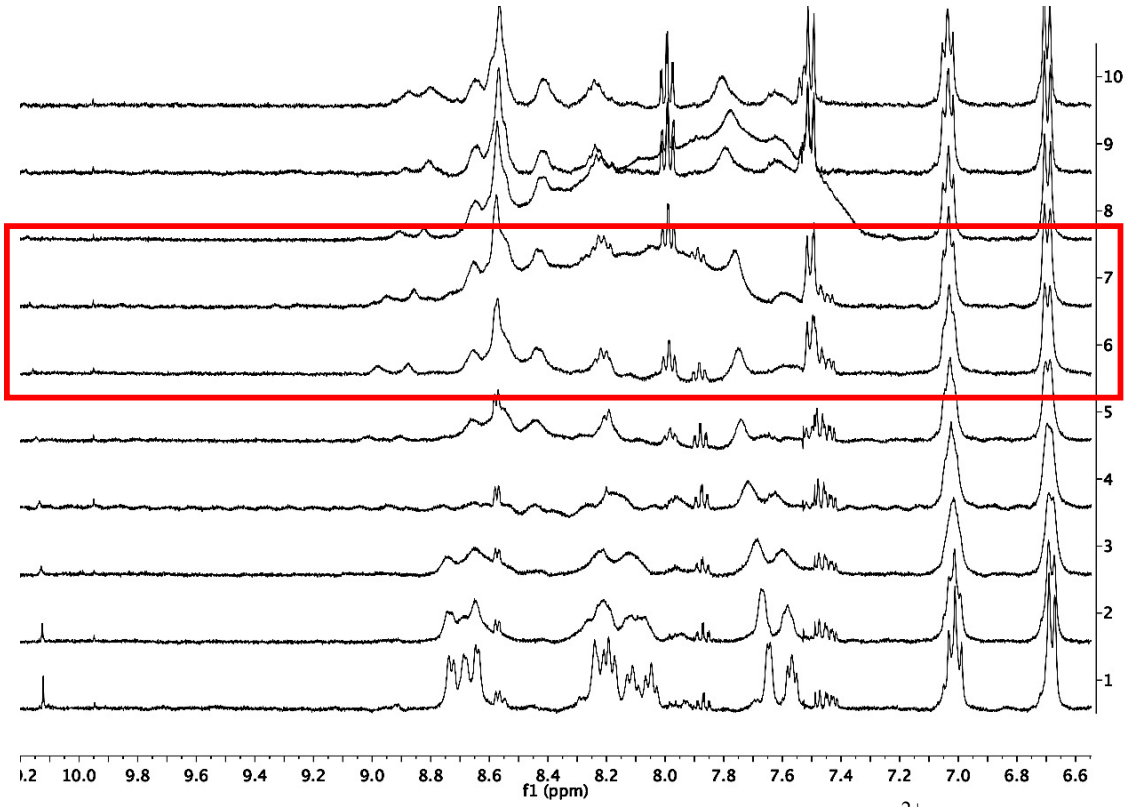
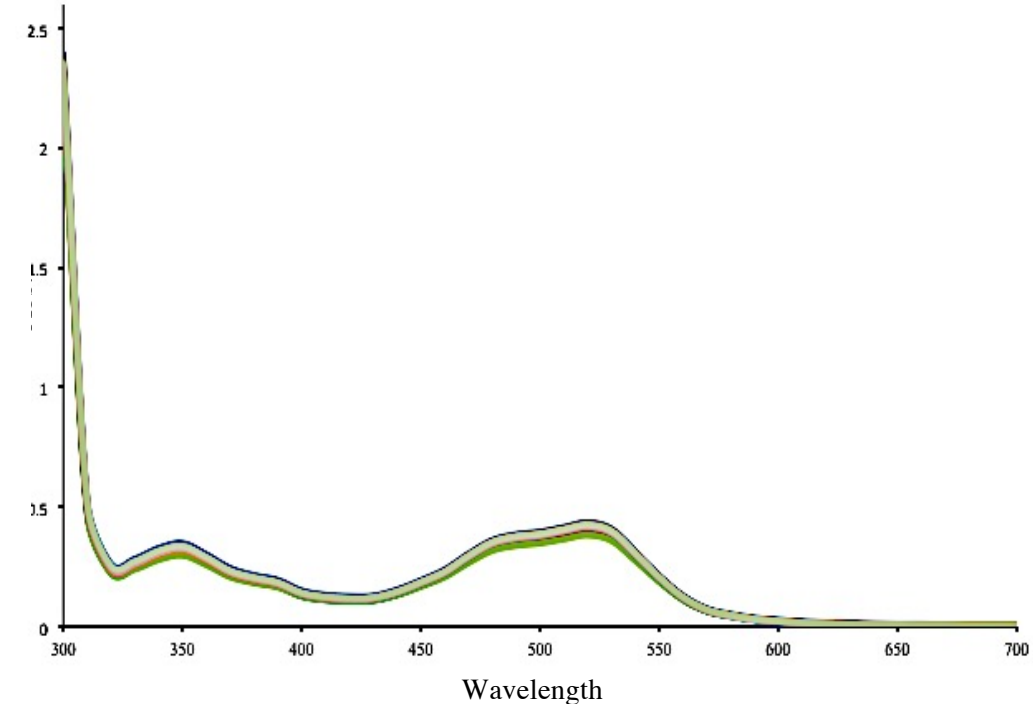


Figure 5.35: Proton NMR titration for MP-5, 2,2'-bipyridine, and Fe^{2+}. Red box indicates 0.833-1 eq. of metal were required for cyclization.



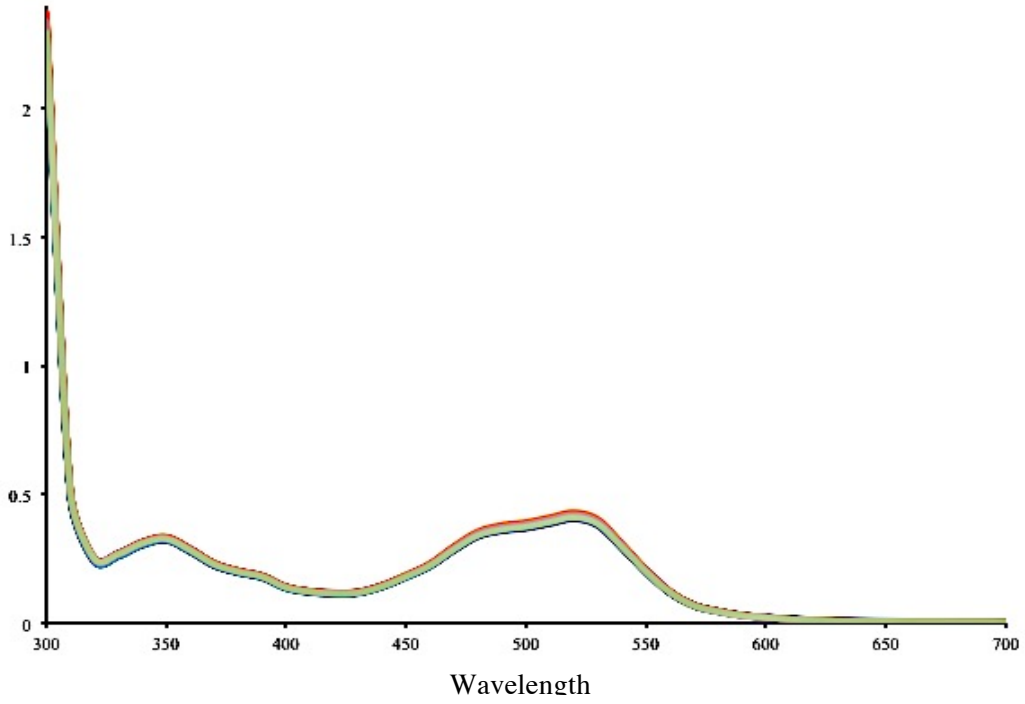


Figure 5.36: (top) Well-plate titration for MP-6, 2,2'-bipyridine, and Fe^{2+} (trial 1). (bottom) Trial 2. Increasing concentration of peptide led to no change in absorbance.

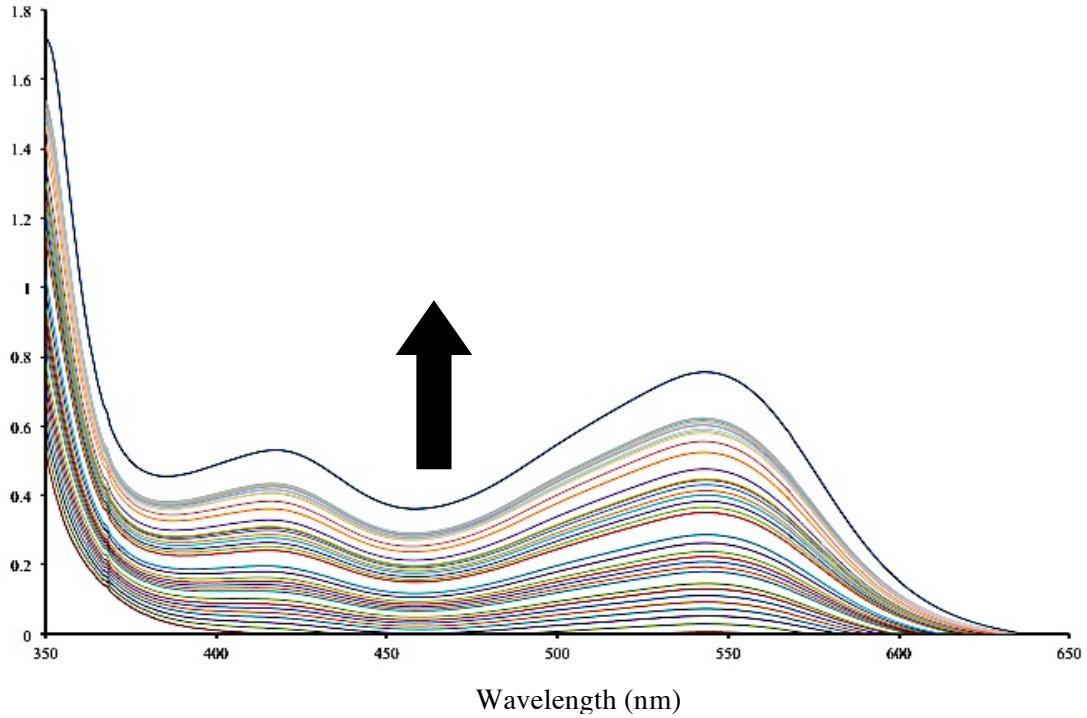


Figure 5.37: UV-Vis titration of MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+}.

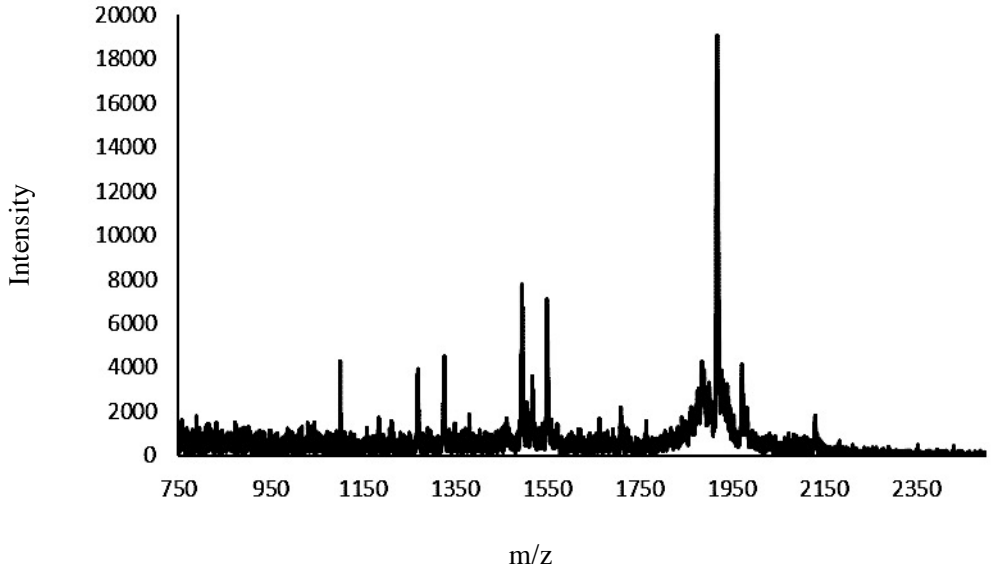


Figure 5.38: MALDI spectrum for MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+}. m/z observed: 1101.02, 1269.04, 1327.06, 1495.11, 1549.58, 1915.

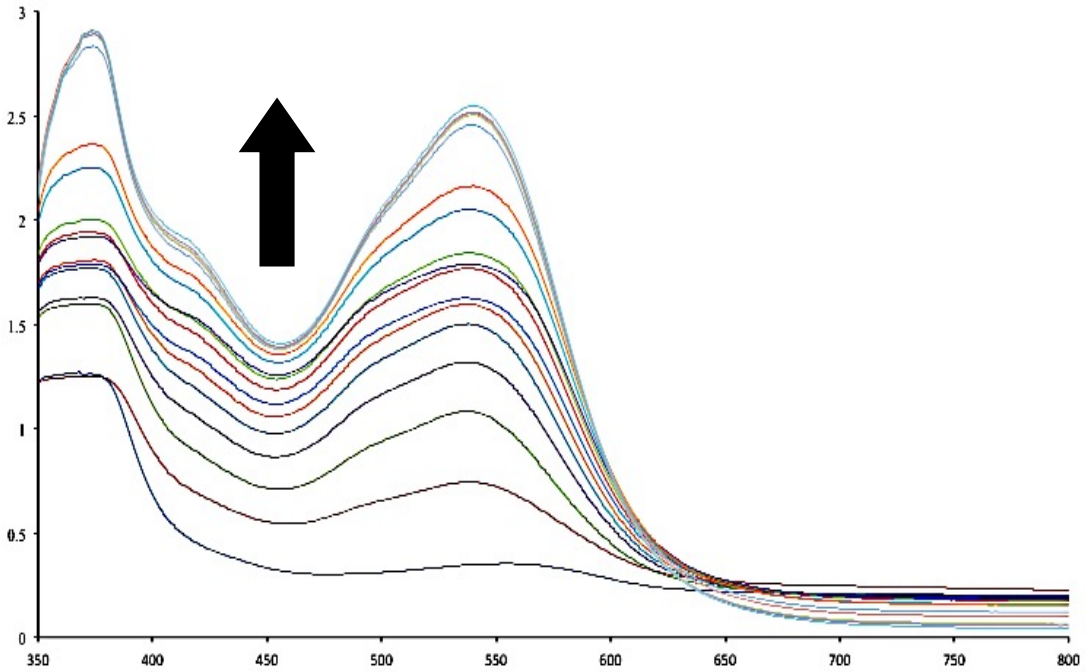


Figure 5.39: UV-Vis titration of MP-5, AldPep/HydPep-2 Bipyridine, and Fe^{2+}.

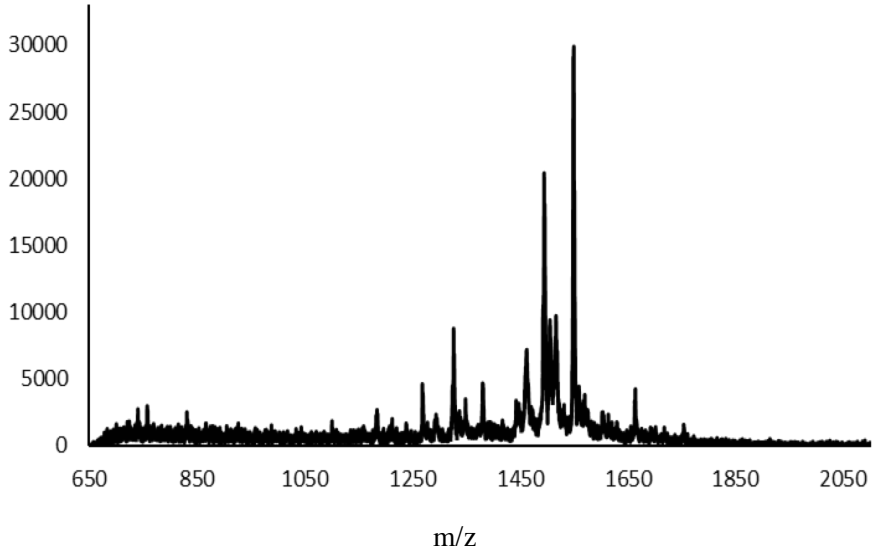


Figure 5.40: MALDI spectrum for MP-5, PepHyd-1 Cyclic Bipyridine, and Fe^{2+}. m/z observed: 1269.10, 1326.91, 1462.06, 1495.17, 1516.91, 1505.39, 1550.06.

5.7 References

1. M. C. Young, E. Liew, J. Ashby, K. E. McCoy and R. J. Hooley, *Chem. Comm.*, 2013, **49**, 6331-6333.
2. E. T. Hernandez, J. Swaminathan, E. M. Marcotte and E. V. Anslyn, *New J. Chem.*, 2017, **41**, 462-469.
3. E. T. Hernandez, I. V. Kolesnichenko, J. F. Reuther and E. V. Anslyn, *New J. Chem.*, 2017, **41**, 126-133.
4. C. Kaes, A. Katz and M. W. Hosseini, *Chem. Rev.*, 2000, **100**, 3553-3590.
5. T. Murashima, S. Tsukiyama, S. Fujii, K. Hayata, H. Sakai, T. Miyazawa and T. Yamada, *Org. Biomol. Chem.*, 2005, **3**, 4060-4064.
6. H. Ishida, M. Kyakuno and S. Oishi, *Pep. Sci.*, 2004, **76**, 69-82.
7. R. Zou, Q. Wang, J. Wu, J. Wu, C. Schmuck and H. Tian, *Chem. Soc. Rev.*, 2015, **44**, 5200-5219.
8. H. Ishida and Y. Inoue, *Pep. Sci.*, 2000, **55**, 469-478.
9. E. A. Ponomarenko, E. V. Poverennaya, E. V. Ilgisonis, M. A. Pyatnitskiy, A. T. Kopylov, V. G. Zgoda, A. V. Lisitsa and A. I. Archakov, *Int. J. Anal. Chem.*, 2016, **2016**, 6.
10. M. A. Zajac, *J. Org. Chem.*, 2008, **73**, 6899-6901.
11. L. Breydo and V. N. Uversky, *Metallomics*, 2011, **3**, 1163-1180.
12. H. M. Seifert, K. Ramirez Trejo and E. V. Anslyn, *J. Am. Chem. Soc.*, 2016, **138**, 10916-10924.
13. K. L. Diehl, J. L. Bachman, B. M. Chapin, R. Edupuganti, P. Rogelio Escamilla, A. M. Gade, E. T. Hernandez, H. H. Jo, A. M. Johnson, I. V. Kolesnichenko, J. Lim, C.-Y. Lin, M. K. Meadows, H. M. Seifert, D. Zamora-Olivares and E. V. Anslyn, in *Synthetic Receptors for Biomolecules: Design Principles and Applications*, The Royal Society of Chemistry, 2015, DOI: 10.1039/9781782622062-00039, pp. 39-85.
14. A. E. Sales, L. Breydo, T. S. Porto, A. L. F. Porto and V. N. Uversky, *RSC Adv.*, 2016, **6**, 42971-42983.
15. W. G. Skene and J.-M. P. Lehn, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 8270-8275.
16. R. J. Wojtecki, M. A. Meador and S. J. Rowan, *Nat. Mater.*, 2011, **10**, 14-27.
17. T. Ganguly, B. B. Kasten, D.-K. Bucar, L. R. MacGillivray, C. E. Berkman and P. D. Benny, *Chem. Comm.*, 2011, **47**, 12846-12848.
18. C.-H. Lu, Y.-F. Lin, J.-J. Lin and C.-S. Yu, *PLOS ONE*, 2012, **7**, e39252.
19. M. M. Yamashita, L. Wesson, G. Eisenman and D. Eisenberg, *Proceedings of the National Academy of Sciences*, 1990, **87**, 5648-5652.
20. J. M. Dragna, G. Pescitelli, L. Tran, V. M. Lynch, E. V. Anslyn and L. Di Bari, *J. Am. Chem. Soc.*, 2012, **134**, 4398-4407.
21. U. McDonnell, M. R. Hicks, M. J. Hannon and A. Rodger, *J. Inorg. Biochem.*, 2008, **102**, 2052-2059.
22. C. J. White and A. K. Yudin, *Nat. Chem.*, 2011, **3**, 509-524.

23. T. Gruendling, S. Weidner, J. Falkenhagen and C. Barner-Kowollik, *Poly. Chem.*, 2010, **1**, 599-617.
24. J. H. van Esch, M. A. M. Hoffmann and R. J. M. Nolte, *J. Org. Chem.*, 1995, **60**, 1599-1610.

Chapter 6: Supramolecular Oligomer Formation of 1-([2,2':6',2''-terpyridin]-4'-ylmethyl)guanidine with Zn^{2+}, Thiophene-2,5-diyldiboronic Acid, and 2,3-Butadione

6.1 INTRODUCTION

6.1.1 Unnatural, Synthetic Oligomer Design from a Peptide Chemist's Perspective

In this final chapter, a new supramolecular oligomer will be discussed. The principles governing the design of this structure originate from peptide chemistry, but added to these principles is the integration of dynamic covalent bonds, as they have been applied to unnatural peptides and unnatural oligomers. These reversible reactions helped facilitate the folding of peptides as discussed in Chapter 5, but first the expertise on how to selectively modify side chains had to be developed. Those selective modification studies served as a foundation for incorporating new residue-specific fluorophores into emerging single-molecule peptide sequencing technologies (Chapters 2 & 3). Selective-peptide modification chemistries also enabled the synthesis of peptides with unnatural residues, such as secondary amine side chains. N-methyl lysine was used for expedient introduction of unnatural functionalities, including boronic acids and bipyridines (Chapter 4 & 5). These groups form part of an effort led by the Anslyn group to use orthogonal, reversible covalent bonds in the synthesis of novel molecules capable of forming higher-order structures.

Peptide bonds rooted the studies described. Their utility demonstrated in nature, they are the backbone of choice for chemists seeking to induce folding of synthetic and unnatural structures.[1] The commercial availability of building blocks, synthetic accessibility, instrumentation, and well-established protocols have made polypeptides a useful scaffold for appending desired functionalities. Usually, these functionalities extend beyond the availability of residues found in nature. For example, the Gehlman group has

extensively studied the use of α, β peptides to improve proteolytic resistance *in vivo* and to probe into the mechanisms of intermembrane processes.[2]

Apart from oligomers with peptide backbones, other types of oligomers, known as foldamers, have been explored for folding studies.[3, 4] These unnatural synthetic structures commonly fold and act as recognition molecules to mimic large molecules like DNA or proteins, which serve as the inspiration governing foldamer design and synthesis. Therefore, many groups have demonstrated formation of helices in their oligomers although other topologies like knots and spirals have been achieved.[5]

Motivation for making these oligomers comes from biomolecule analogues capable of matching or exceeding a particular biological function. Usually, a common goal is to ensure a robust compound not prone to degrading. Therefore, a synthetic structure that can fold like protein or DNA and that is also resistant to degradation inside an organism could be of great biological value, thus the interest in foldamer research. Synthetic, unnatural structures that can resist degradation, remain as inert entities, and perform only their intended purpose remain a focus of research.[6] Functions can include but are not limited to binding, or molecular recognition. From these binding studies, foldamers have also led chemists to study unique motions such as rotations and screwing motions.[7]

6.1.2 Limitations of Unnatural, Oligomer Design and Synthesis

A current limitation when considering making biomimetic oligomer structures is the synthetic difficulty of the building blocks, especially when incorporating side chain functional groups. If these unnatural oligomers are to begin to resemble biomolecules like proteins, appending functionalities off the backbone should become part of the synthetic design. Many of the examples in the literature show building blocks that require an extensive and skillful execution of synthetic steps that deter from the intended goal of analyzing folding/binding properties. In addition, these building blocks tend to lack side

chain functionalities that expand the diversity of monomers that can be made from a set of common chemistries. Complicated building block synthesis is a bottleneck, but the oligomerization, or elongation, process is also a stumbling block. Oligomerization of peptides is not atom economical and produces large amounts of waste, but it remains widely used because of the automated instrumentation and commercial availability of building blocks and reagents. The prevalence of the amide bond in nature has also easily justified the synthesis of peptides, despite their costly and wasteful production. The same cannot be said for the unnatural oligomers that are conceived, produced, and studied by one or a few groups.

As a result, a facile means of producing, screening, and scaling up the synthesis of an application-worthy unnatural oligomer continues to be a slow process. Even when using well-established peptide chemistries, as described in Chapter 5, appending and identifying unnatural functionalities which are useful in a material or biological setting still present challenges. For example, designing expedient means of modifying residues and scaling up the synthesis of building blocks with these non-canonical residues can be a time-consuming process and an inefficient use of resources for peptides. This becomes a process that incurs considerable risk especially at the preliminary stages of proving the utility of an unnatural residue. Chapters 3 & 4 explored straightforward ways to modify peptides, using known building blocks, or solution-phase chemistries, that facilitated introduction of secondary amines. These secondary-amine peptides could be considered as simple screening peptide scaffolds that can be modified by alkylation with bromide compounds. If an unnatural functional group gave a positive result, the building block with that residue could be readily scaled up. The development of such a methodology was possible because of the extensive maturation of the area of synthetic peptide chemistry. Similar access to chemistries to modify unnatural oligomers for screening purposes is virtually non-existent.

6.1.3 Lessons from Peptide Chemistry that can be Applied to Unnatural Oligomers

The first lesson that can be taken from peptide chemistry is to have synthetically accessible, structurally simple, and easy to modify building blocks. The twenty amino acid residues are either commercially available or relatively synthetically accessible for solid-phase synthesis. Unnatural oligomers should preferably be made of building blocks acquired commercially or easily scaled up in the lab. If not, it can limit the diversity of structures that can be screened in a timely fashion. Structurally these compounds should be simple in a way that the diversity of the compound originates from the side chains branching off the backbone. That the building block itself must serve as a scaffold underpins the design of an oligomer's backbone. It should be readily modified to induce unique folding structures. The natural analogues are the residues in proteins that can hydrogen bond, aggregate due to hydrophobic effect, or participate in π-π interactions.[8] Unnatural building blocks with these moieties branching off the backbone should also promote folding.

The second lesson is to use a reliable "linking" chemistry. A host of amide coupling reagents, protocols, solution-phase, and solid-phase methods optimized for dependable polypeptide formation exist.[9, 10] The elongation process must occur consistently and efficiently, ensuring the quality of the target sequences. Avoiding deletion products is at the root of these optimization efforts. Similarly, for unnatural oligomers to be made and reliably screened, there must be approaches that ensure unnatural oligomers are made with few deletions.

Lastly, it is important to form an oligomer that can be modified at a side chain along the backbone that can serve as a branching point from which more functionality can be introduced. In addition to building an alphabet of building blocks that share a common simple structure differing in side chains, another set of building blocks should be readily

modified, post-oligomer synthesis, especially if making a synthetically challenging or highly reactive building block. If the functionality to be added to the oligomer cannot survive the elongation process, it precludes the ability to make a building block. Having functionalities that can react post-oligomer synthesis provides a convenient entry to bring about a diverse set of oligomers from a common origin.

6.1.4 Modular Oligomer Design

In a broad sense, the three principles discussed in 6.1.3 can be subsumed under the term of modular oligomer design. This term will be used when one of these three principles is addressed throughout the text. Modular oligomer design will also allude to the focus of reducing complex higher-order structuring of unnatural oligomers into simple, intellectually digestible forms based on the straightforward design and accessibility of the components required for oligomer formation.

This work does not argue that the new oligomer proposed here necessarily marks an improvement over unnatural oligomers in the literature. Nor, does it claim that our oligomer surpasses the form, function, and utility of polypeptides. The goal of this work focuses on applying modular oligomer design to make an oligomer from building blocks that are commercially available/or synthetically accessible, are linked by a simple reliable chemistry, and contain chains that are easily modified post-oligomer formation. More importantly, focusing on re-conceptualizing the design and synthesis of foldamers and building upon the disciplines encompassing supramolecular, dynamic, combinatorial, polymer, and biological chemistry are the goals of this work. As in all these fields, chemistry presented here takes inspiration from and strives to emulate what exists in nature.

6.1.5 A Guanidinium-Terpyridine Supramolecular Oligomer

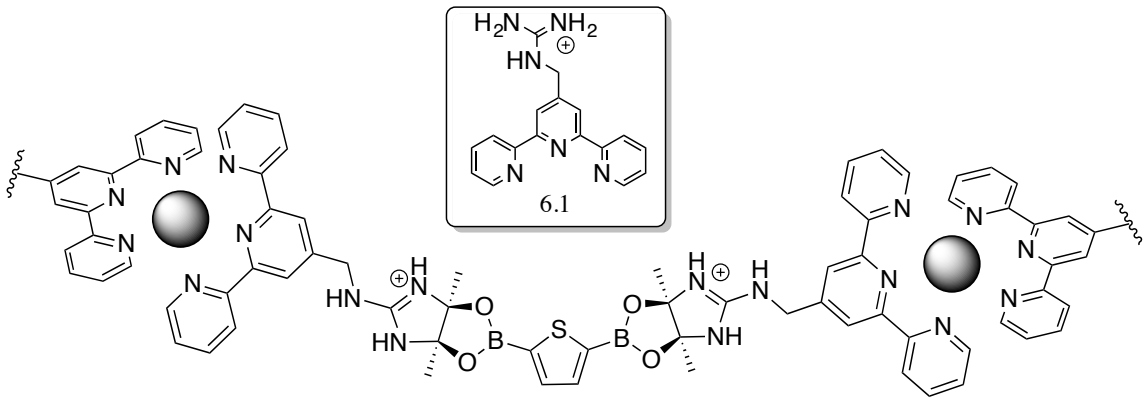
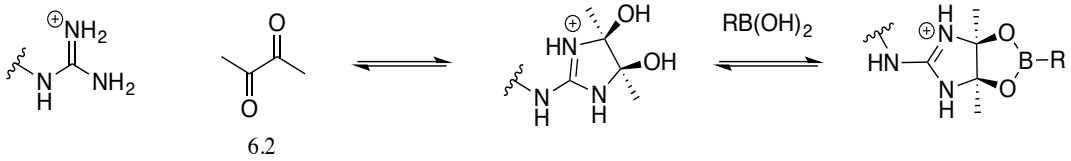


Figure 6.1: Proposed Oligomer Using Guanidinyl Terpyridine, Zn^{2+}, Di-boronic acid, and Butadione.

Figure 6.1 is the generalized scheme envisioned for inducing oligomer formation of 1-([2,2':6',2''-terpyridin]-4'-ylmethyl)guanidine (compound 6.1). The building block is formed by the reaction of the adjoined guanidinium terpyridines , as shown in equation 6.1.

eq. 6.1



6.2

The guanidinium group, at pH values greater than 4,(? Why 4?) can readily attack diketones like 6.2, forming a five-membered vicinal diol ring. If a boronic acid is also present, the five membered diol ring binds to the boron center, such that the dione bridges the boronic acid and the guanidinium. Such an interaction was developed in the 1970s and has been utilized to inhibit protein function in enzymatic mechanistic studies.[11] Mass spectrometry has reliably exploited this reaction for selective and global modification of arginine

residues as mass labels, helping resolve peptides with similar molecular weights.[12] This reaction also enriches samples with arginine-rich peptides;[13] solid supports with immobilized boronic acids are incubated in a peptide digest and butadione. Peptides with guanidinyl residues bind to the solid support. Those without remain in solution and are washed away from the solid support with buffer and butadione. Finally, incubating the resin at pH 4 reverses its reaction with arginine and dione, and so bound peptides come off the resin.

The pH-based reversibility of this three-component assembly makes it attractive in the proposed oligomer in Figure 6.1, which could be decomposed to its individual components by adjusting the acidity of the solution. The formation of the oligomer could be viewed as "clicking", and its decomposition as "declicking", with applications similar to conjugate acceptor declicking reported by the Anslyn group.[14]

The oligomerization and polymerization of terpyridines has been extensively studied.[15] The well-known and robust metal-ligand complexation was therefore selected for modular oligomer design as the analogue to the amide bond in polypeptides (Figure 6.2). The bicyclic junction resulting from the assembly of a dione, guanidinium, and boronic acid, from which side chains can emanate, can be considered the analogue to side chains that branch off the peptide backbone. Figure 6.2 demonstrates the side chains coming off the dione, and greater complexity can be added by using asymmetric diones. There are several forms of diversification that can be achieved by using an asymmetric dione. The side chain would be oriented in a randomized manner, but possibly stereoselective reactions could be devised to consistently direct the orientation of the side chain. A mixture of symmetric or asymmetric compounds would form complex mixtures of sequences in a straightforward fashion. By merely mixing simple components that are

either commercially available or easily synthetically accessible, a quite complex mixture of products may be achieved.

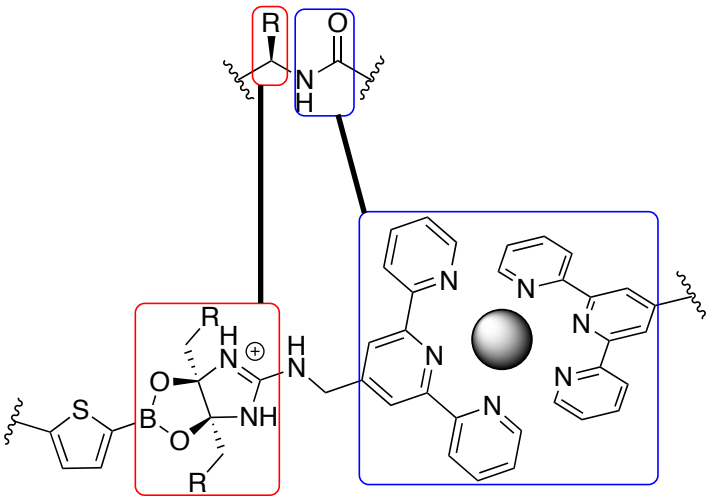


Figure 6.2: Comparing Peptide Backbone to Proposed Oligomer Backbone.

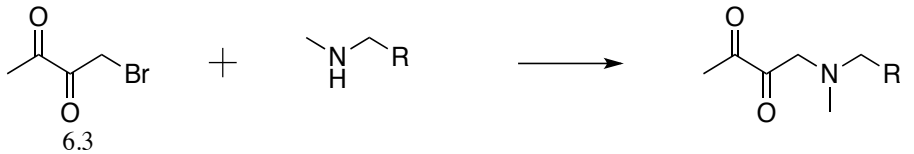
High affinity metal complexation by two terpyridines is the dependable chemistry, much akin to amide bonds, that provides a linker allowing for elongation of the oligomer. This linker was chosen based on previous work wherein oligomers were formed from terpyridine capped pegylated chains.[16] However, the flexible pegylated chains caused the oligomer to be prone to cyclization. The design of oligomers resembling biomolecules requires semi-rigid backbones.[4] Flexible homo-oligomers tend to form globules that lack a defined and rigid structure.[4] Our linking strategy therefore incorporates moieties that are more constrained at the branching point.

The branching point is unique in this oligomer because it uses supramolecular linking chemistry. Without this three-component assembly between the guanidinium, dione, and boronic acid, metal complexation could not result in oligomerization, and *vice versa*. Therefore, a two linker-dependent structure arises. This feature offers a particular

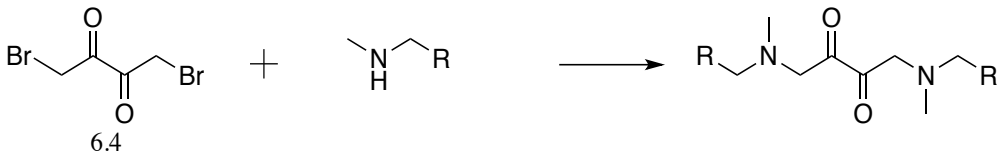
advantage for promoting oligomer formation or degradation by at more than one type of junction. Without the metal, the oligomer does not form at the terpyridines. Without the butadione, the oligomer cannot form at the guanidiniums and boronic acid junction.

Finally, the backbone follows the principle of oligomer modular design, because it has branching points capable of introducing functionalities that can be easily modified by other chemistries. Therefore, we propose our oligomer will serve as a convenient scaffold for the formation of a diverse array of supramolecular oligomer structures. The commercially available bromobutadione can be modified by alkylation, as shown in eqns. 6.2 and 6.3.[17] More complex diones may be synthesized from diverse structures according to literature.[18]

eq. 6.2



eq. 6.3

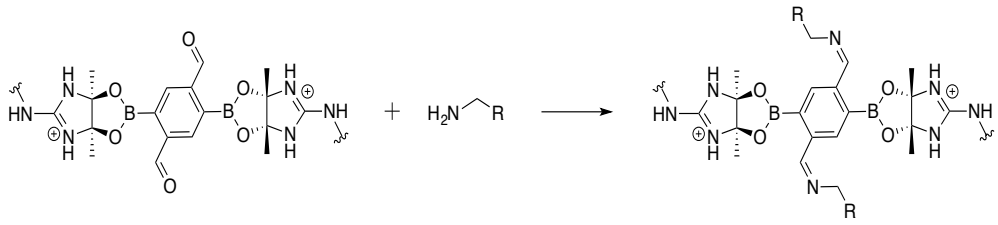


As long as functional groups are stable to alkylation conditions, we propose this will be a facile means for derivatizing butadiones and accessing a wide set of side chain functionalities.

Another branching point for the oligomer derives from the boronic acid. Diformyl di-boronic acid compounds could serve as a different means for introducing more

functionality (equation. 6.4). Imine formation would be a viable way to expand side chains functionalities and their attachment could be made permanent by reducing the imines to amines. More importantly, the commercial availability of di-boronic acids possibilities the formation of novel structures that could arrange in different forms depending on the positioning of boronic acids on a small molecule, for example if they are positioned *ortho*, *meta*, or *para* to each other. This feature adds to the possible diverse structures envisioned to form with the oligomer. Furthermore, non-covalent functionalities known to polymerize can be appended to these branching points, which can cross link the proposed oligomer. All in all, according to the principles stipulated in previous sections, we claim the proposed oligomer has been formulated considering modular oligomer design.

eq. 6.4



6.1.7 Possible Improvements When Compared to Peptide Chemistry

One advantage the guanidinium-terpyridine oligomer structure may have over peptide chemistry is the improved atom economy. Foremost, solution-phase chemistries underpin the elongation process. Unlike solid-phase peptide synthesis (SPPS), which requires excess amount of starting material to diffuse throughout the solid support,[19] our approach moves toward stoichiometric amounts. Therefore, cost and waste of material may be improved. Furthermore, by lowering the pH below four, the oligomer can decompose

to its original components, so that the starting material may be theoretically recovered for future use. The only material used in excess is the dione, since excess is required for oligomers to form. However, the relative inexpensiveness of diones partially mitigates this drawback.

Although the number of units that can come together to form this oligomer remains to be determined, the number may possibly exceed that of SPPS. Currently, state-of-the-art technologies push the limit of amino acid residues up to 40.[20] Although impressive, that amount can be exceeded by the supramolecular interaction in metal complexation. Large molecular weights have been reported for terpyridines polymers, which provide literature precedent for linking a considerable number of building blocks in our oligomer synthesis.

Methods do exist to make high-molecular weight polyamides; however they tend to lack the ability to incorporate several side chain functionalities.[20] Furthermore, if different side chain functionalities were to be added, a particular sequence would be difficult to replicate, due to the non-reversibility of the amide bonds. Thus, for polypeptides with diverse side chains, SPPS remains the premier manner to reproduce desired sequences.

.

6.1.8 Dynamic Covalent Bonds to Overcome Issues with Sequence Reproducibility

A similar argument can be made for our oligomer. How can we reproduce a desired sequence for a particular application? At the root of this work, is the exploitation of the dynamic covalent bonds described in Chapter 5. Much focus has been directed towards these reactions in the formation of novel higher-order structures.[21] The ability for reactions like hydrazone formation, boron esterification, and thiol conjugate addition to form products but readily reverse, coupled with metal-ligand complexation of terpyridines, makes for attractive intermolecular interactions. They serve as an intermediate between the

rigid, highly irreversible covalent bond and the highly dynamic non-covalent bond. In other words, these dynamic covalent bonds are regarded as more stable, but prone to change in the presence of competing reactants with similar functional groups. Thus, manipulating the kinds of reactants and their proportions in dynamic covalent systems offers a degree of control over synthetic structures chemists seek.

These reactions become a directional driving force that allows for reliable means to form higher order structuring.[4] Non-directional forces like those of the hydrophobic effect and van der Waals interactions can still be incorporated, along with directional ones such as hydrogen bonding. These dynamic, covalent bonds add to the tools available for improving folding of synthetic structures.

In the case of inducing folding based on the molecular recognition of a target analyte or a set of analytes, we hypothesize the oligomer described should display different kinds of sequences and structures. What governs the elongation process is not a covalent one, but a supramolecular one. Combined with the dynamic reactions mentioned, this oligomer should not be limited by the need to reproduce a target sequence, because inherent in this system will be the process where the oligomer will find its thermodynamically favored state; which can change in sequence and structure depending on the environment. Therefore, the means to control the sequence of the oligomer will not depend on the elongation process similar to SPPS, where building blocks are incorporated by a linear, step-wise process, but instead by controlling external conditions of solvent, reactant proportions, and analytes present in solution.

6.1.9 The Dynamic/Orthogonal Nature of the Guanidinium-Dione-Boronic Acid Interaction

An important feature stressed here is the presumed dynamic nature of the reaction described in equation 6.1. From Chapter 5, we know that boronic acids are dynamic and

reactants with similar starting materials readily exchange. Part of the tri-component ensemble is boronic acid esterification and is therefore regarded as dynamic. The attack of the guanidinium on the dione is reversible, since binding by a boronic aid is needed to lock in the five membered ring. Currently, research in the Anslyn group focuses to confirm the dynamic nature of equation 6.1. If shown as dynamic, this tri-ensemble will be one of the first, to our knowledge, to be reported. Furthermore, the reaction between guanidinium and dione will be another reaction considered as dynamic that, to our knowledge, remains unreported.

As far as the orthogonality of this tri-component assembly, boronic acid binding to diols does not cross react with amines, thiols, terpyridines, and α,β unsaturated ketones.[22] Therefore, the boronic acid binding event will be quite inert. However, a little more difficult will be to assess the orthogonality of diones. At least when considering the dynamic reactions previously reported, diones should have little to no cross reactivity. Amines may condense readily, however, in aqueous conditions condensation reverses. Since, these diones may be selective under wet conditions, efforts will be aimed to form this oligomer in aqueous environments. Thiols may also condense; so further studies to determine the reversibility of these products in water are important.

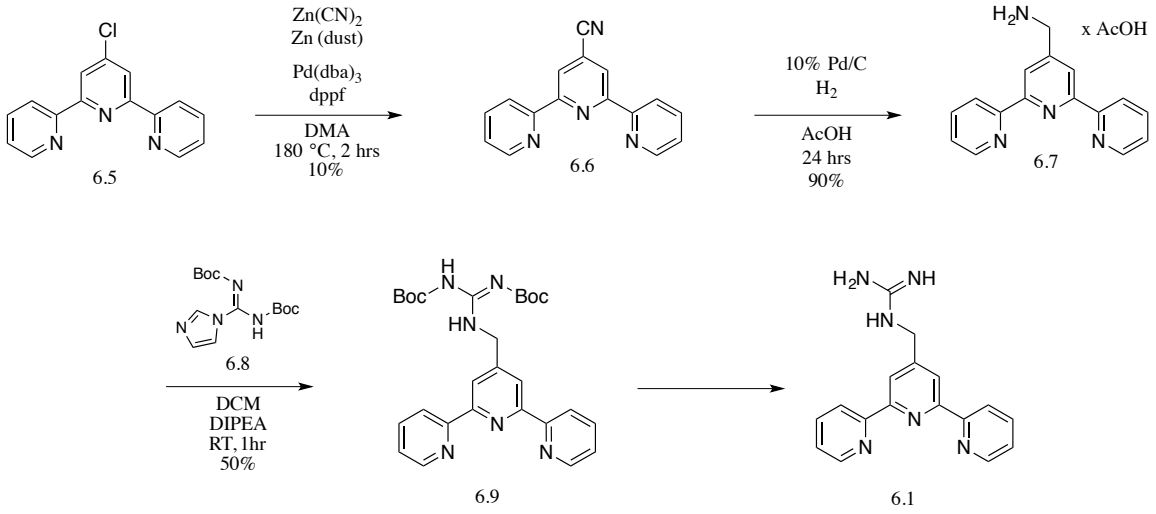
6.1.10 Scientific Aim: Order of Addition and Oligomer Formation Studies

In this chapter, experiments performed aimed to answer two questions. Does it matter in what order the components were added to form the oligomer? In other words, we wanted to explore if any cross reactivity within the tri-component assembly arose by doing a combination of different order of additions. Also, does an oligomer form when a terpyridine, metal, and the tri-component ensemble are present in an aqueous solution? To answer these two questions, proton NMR studies were carried out to determine if order of addition mattered. An extension of this study was determining how much dione promoted

assembly formation. Boron NMR became the method of choice to monitor the equivalence needed. Proton DOSY NMR was also used to try to establish the formation of oligomers from observed changes in diffusion that suggested formation of large molecules. Studying the oligomerization is the first step towards probing the presumed emergent properties described in previous sections.

6.2 Results and Discussion

6.2.1 Synthesis of Compound 6.1



Scheme 6.1: Synthesis of Compound 6.1.

Scheme 6.1 shows the synthesis of 6.1. Compound 6.5 was commercially purchased and converted to 6.6 following literature protocol. Reduction of this cyano derivative afforded compound 6.7, which was guanidinated with 6.8, also commercially available. Quantitative deprotection occurred with 20% v/v TFA in DCM. Leaving the compound in salt form facilitated dissolving guanidinyl terpyridine into aqueous mixtures. These ligands normally do not dissolve in water, and incorporation of pegylated chains helped improve solubility.[23] Compound 6.1, to our knowledge, is not reported in the

literature, and can serve as an alternative to pegylated derivatives for solubilizing terpyridines in water.

6.2.2 Order of Addition Studies

The boronic acid, guanidinium, and butadione react with each other to form the tri-component assembly. However, the reactivity of compound 6.1 with any of these compounds had not been established. and so order-of-addition studies were performed. In Figure 6.3, the different permutations along with the compounds used in the NMR experiments are listed. Compound 6.11, denoted "B" in this section, was provided

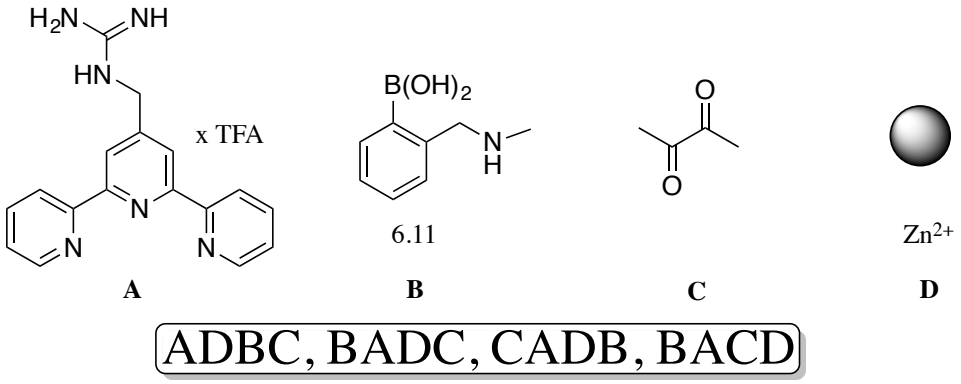


Figure 6.3: Components Used in Order of Addition Studies and Orders Studied.

by the Anslyn group. The improved binding of diols at neutral pH with ortho-(aminomethyl) boronic acids was described in Chapter 4. This compound served as the simplest model available for boronic acids with ortho benzylic amines with minimal steric hindrance. Butadione was chosen for the same minimization of steric hindranceZn^{2+} is bound strongly by terpyridines and is compatible with NMR spectroscopy.. Fe^{2+} is also strongly bound by terpyridines, but its paramagnetic properties would broaden signals. Broadening or loss of fine structure was to be interpreted as part of assembly formation, and so paramagnetic iron could give false positives.

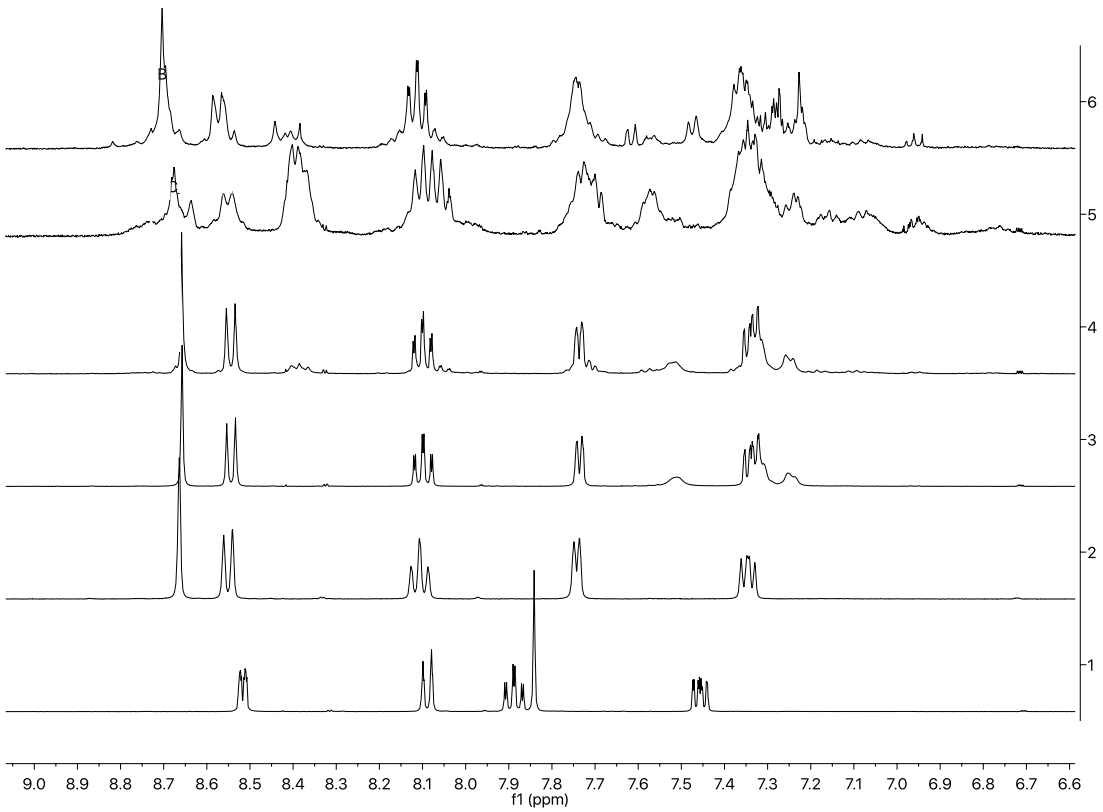


Figure 6.4: Proton NMR Studies for A, D, B, C Order of Addition.

A, D, B, C was the first order studied (Figure 6.4). The stoichiometry between the terpyridine and metal (A and D) required careful control, since excess zinc caused mixtures of mono and fully bound ligand. Binding the terpyridine ensured the pyridine rings did not cross react with butadione. Broadening of the boronic aromatic signals occurred when added the boronic acid was added to the fully bound ligand (ADB). This loss in fine structure only happened for the boronic acid and not the terpyridine rings, suggesting the possibility of the guanidinium interacting with the aromatic portion of the boronic acid. This broadened signal persisted after introducing one equivalent of butadiene (ADBC). Determining if assembly occurred remained dubious, because signals still retained fine

structure. When a total of ten equivalents butadione were added, signals did broaden, but no peak could be identified as indicative of assembly forming. To see if further change occurred by adding more butadione, a total of 100 equivalents were added. Interestingly, the peaks did not further broaden, but instead fine structure for signals returned. A cluster of new peaks formed between 6.9 and 7.3 ppm, but remained unidentified.

Although a peak identifying assembly formation was not apparent in the spectra, studies continued. The rationale for continuing derived from the idea that if order did not matter, whether a peak proving assembly was found or not, the final spectra of all species combined should look the same. Thus, the next order (B, A, D, C) had the boronic acid as its starting point. The aromatic region for B again broadened, albeit with metal-free terpyridine ligand. As with bound ligand, unbound terpyridine still retained fine structure in the aromatic region. Adding zinc did not change the boronic acid, but complexation occurred readily. After adding 1 equivalent of dione C, signals broadened with the surging of peaks as described earlier. The final spectra looked the same as for A, D, B, C. Instead of adding more equivalents of butadione, the solution sat RT overnight. The following day the spectrum did not change, suggesting that equilibrium was had been reached within an hour of dione addition.

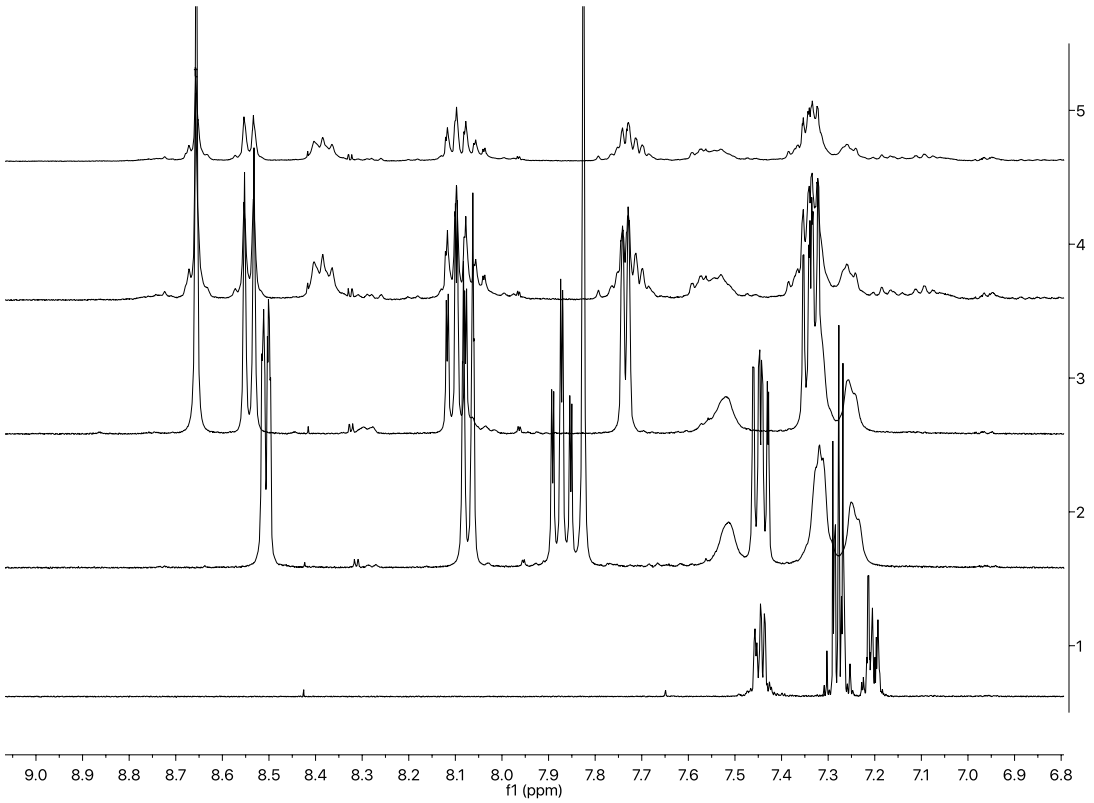


Figure 6.5: Proton NMR Studies for B, A, D, C Order of Addition.

. As with the previous NMR study, broadening of peaks did not confirm assembly formation, but acquiring a final spectrum that did not differ from the one before suggested no dependence on addition order.

A third combination (C, A, D, B) aimed to determine if reactivity between pyridine and butadione existed. In the aromatic region, nothing was observed as expected. After adding A, some impurities were observed (Figure 6.6). Butadione could have reacted with the terpyridine rings to give a mixture of unintended byproducts.. Adding zinc did not reverse the impurity peaks.

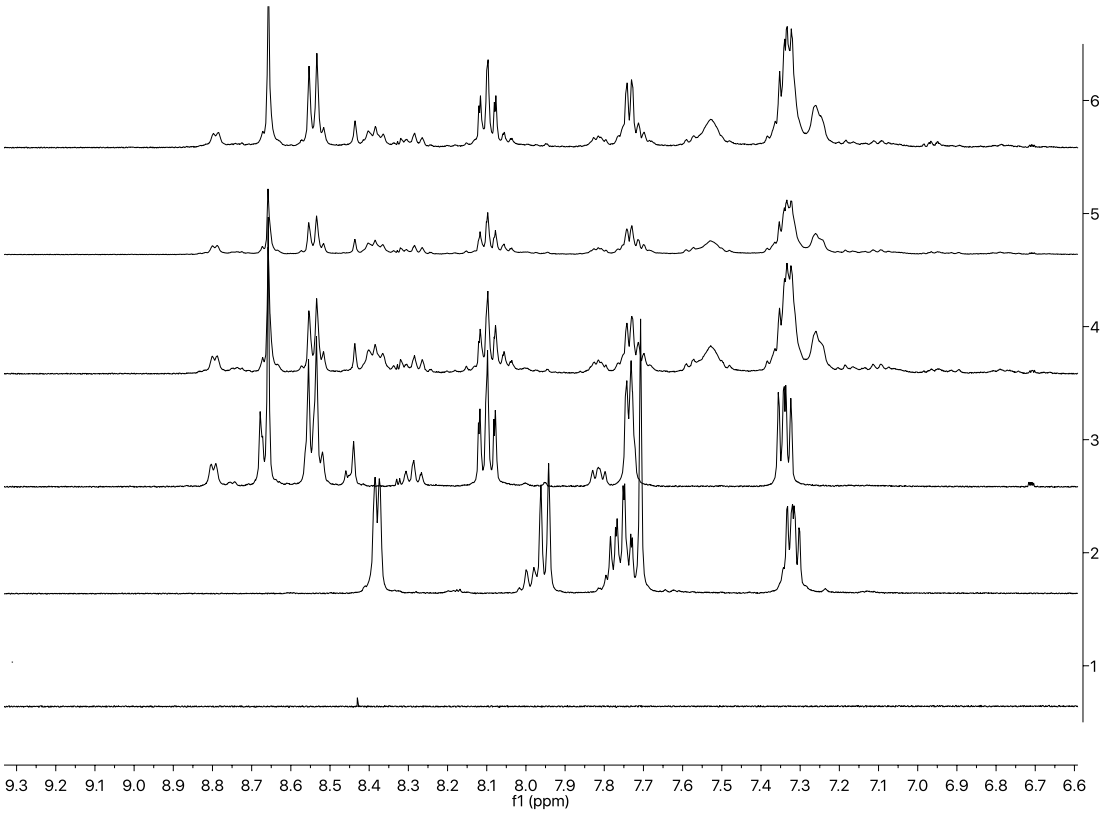


Figure 6.6: Proton NMR Studies for C, A, D, B Order of Addition.

We interpret that the order of addition in this particular case led to unintended side products. However, an alternative hypothesis is that these impurities come from non-zinc metals, such as Fe^{2+}, picked up from the lab (spatulas, glassware, etc.)when handling compound 6.1 over time. However, further studies to test the hypothesis were not pursued. Aside from these impurities, the spectrum for CADB looked like those taken previously.

This sample was heated at 37 °C and 50 °C to see if the spectrum changed, but no differences were observed. However, some cooling occurred prior to taking an NMR, and there could have been a reversion to the same product mixture that existed at room temperature prior to heating. We propose to study temperature-dependent NMR in the future to directly observe the effects of heating.

Finally, order D, A, C, B (SI) completed the set. As with all three previous NMR studies, the final spectrum looked the same. Thus, despite the slight impurities observed for C, A, D, B, order did not play a large role in assembly formation. Components could be mixed together simultaneously with minimal concerns for side products.

6.2.3 Boron NMR Titration Studies

Since proton NMR studies alone failed to unambiguously confirm assembly formation, boron NMR studies became an alternative. Second to the carbonyl carbons, the boron should experience the greatest change between its unbound and bound state. Therefore a titration was performed while monitoring boron to determine the equivalents of dione required for assembly in buffered aqueous mixtures. Literature accounts report the excess use of butadione to ensure efficiency without biasing results.[12] However, if improving the atom economy of an unnatural oligomer formation were the aim, determining the minimum amount of butadione required would be important. We presumed that excess was not required, because of butadione's reactivity. However, water could outcompete the resulting diol for binding by the boronic acid. Entropy should disfavor diol binding. In contrast, enthalpy should favor diol binding.

Titration studies helped to settle these questions. Initially, the components in section 6.2.2 were used. Interestingly, the boron NMR signal disappeared after introducing the metal-ligand complex. Control spectra of compound 6.11 with only Zn^{2+} or only 6.1 did not cause signal loss. The metal bound ligand complex broadened the signal to such an extent that it was not discernible from the baseline.. Upon the addition of one equivalent of butadione, a new peak emerged at 9 ppm, which we interpreted as assembly formation. At 5 equivalents, the signal intensity once more decreased to the baseline. With 30 equivalents butadione, a peak at 19.5 ppm became more prominent. Dilution of the sample with each addition of butadione certainly contributed to loss in signal intensity, but a

dilution factor of 1.7 could not solely account for the disappearance of signal. Thus, it became difficult to assess the minimum amount of butadione required for complete assembly formation.

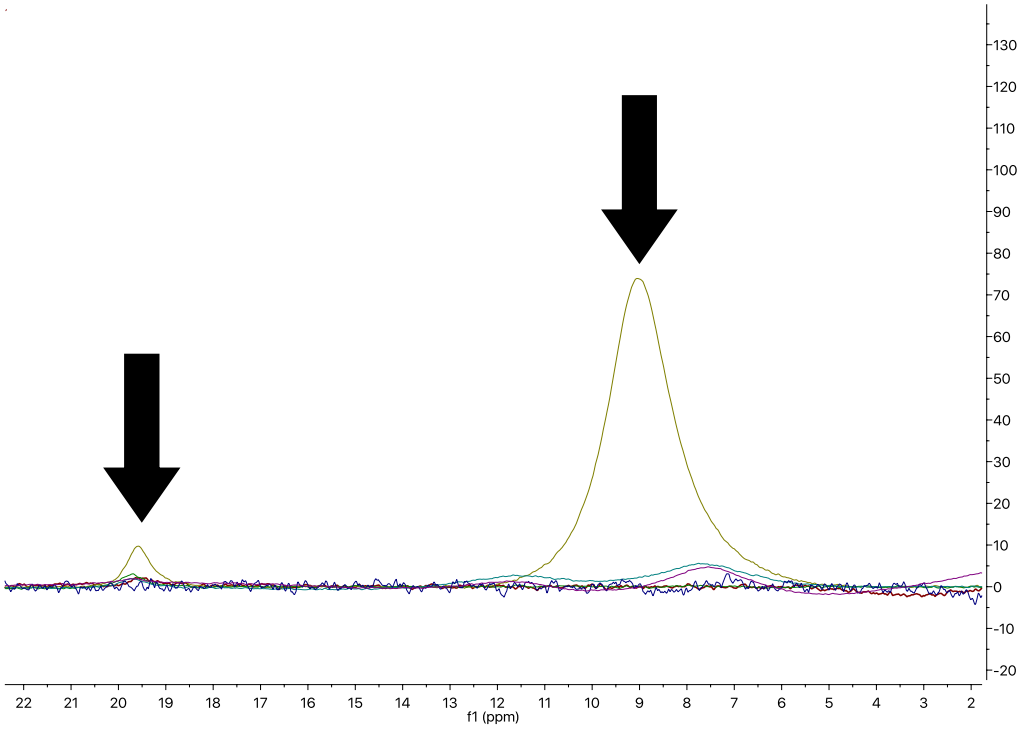


Figure 6.7: Boron NMR Titration of Butadione into a Solution of Metal-Bound 6.1 and 6.11.

With increasing excess of butadione in solution, we propose that reaction of the secondary amine in the *ortho*-aminomethyl phenylboronic acid with butadione became more prevalent, disrupting the pseudo tetrahedral, or sp^3, arrangement of the boronic acid.[24] Chemical shifts between 19-20 indicate sp^2 formation, not consistent with *ortho*-aminomethyl phenylboronic acids binding to vicinal diols. In other words, compound 6.11 was not useful in this type of experiment because of the secondary amine. A tertiary amine, such as that in compound 6.12, could react with butadione, but the ammonium product would readily reverse in aqueous systems. (Figure 6.8).

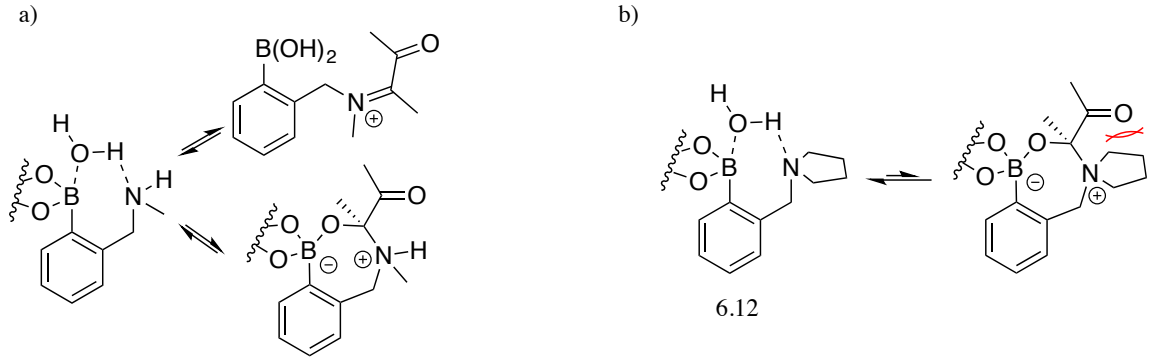


Figure 6.8: Reactions of Compounds 6.11 and 6.12 with Butadione. a) Secondary amine would attack the carbonyl carbon to form an iminium (resulting in an sp^2 boronate) or an alkoxide that can esterify the boronic acid (resulting in an sp^3 boronate). b) Tertiary amine cannot form the imminium, only the alcohol, but it is less prone to esterify to the resulting sp^3 boronate due to steric repulsion between the five-membered ring and the carbonyl.

In fact, if the secondary amine in compound 6.11 attacked butadione as shown in Figure 6.8a, the broadening of the boron signal could be rationalized by unintended oligomerization or polymerization as another aminomethyl phenylboronic acid attacked the remaining carbonyl as neighboring boronic acids partially esterified. Thus, a vast mixture of products formed making the boron signal so broad and difficult to identify what occurred. In contrast, an attack by compound 6.12 on butadione should readily reverse, due to steric issues rising from the close interaction between the five-membered ring and the ketone adjacent to the alcohol formed (Figure 6.8b).

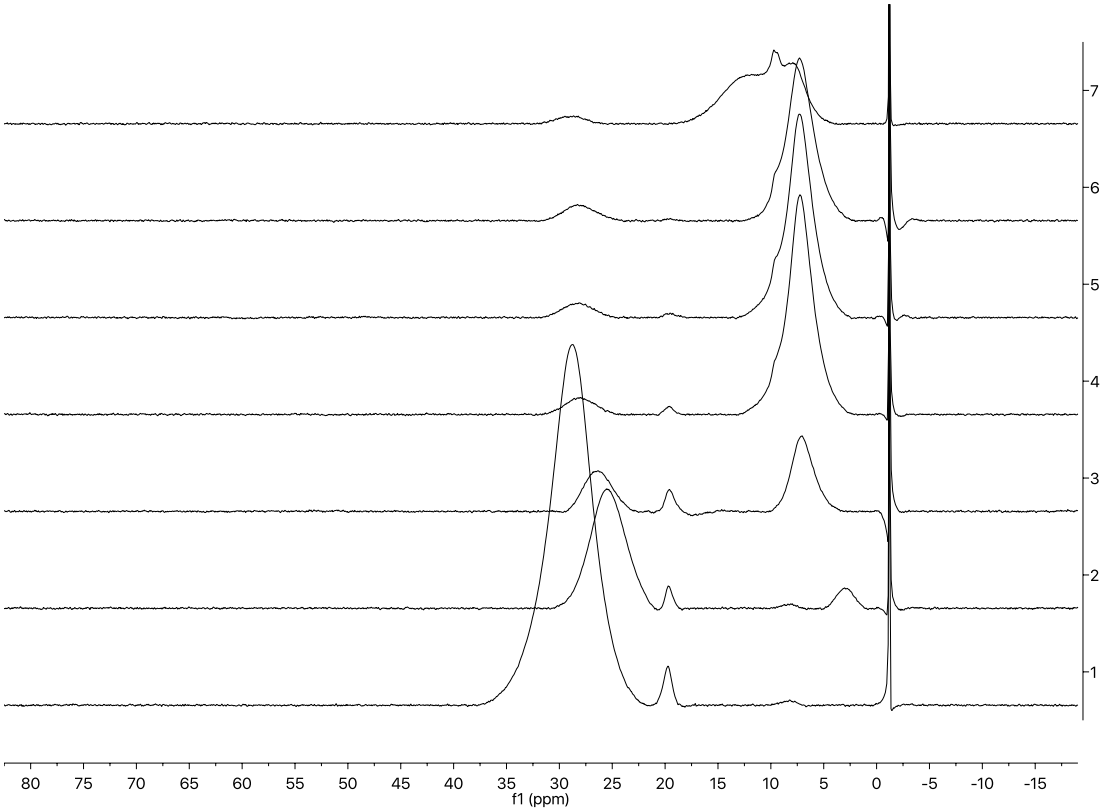


Figure 6.9: Boron NMR Titration of Butadione into a Solution of Metal-Bound 6.1 and 6.12.

The titration with compound 6.12 provided a more consistent change from a trigonal planar to a tetrahedral arrangement, consistent with boron binding to vicinal diols. Between 7 and 10 equivalents the NMR stopped changing considerably. Therefore, the minimal amount required for assembly formation in neutral conditions was set at 10 equivalents. Although not ideal when improving atom economy and decreasing waste in oligomer production, ten equivalents was deemed acceptable. Interestingly, at 30 equivalents, aside from dilution causing a decrease in signal intensity, the singlet peak at 7.5 ppm broadened. The residual peak at 30 ppm, attributed to unreacted boronic acid, was not completely eliminated though. In addition, along with broadening some fine structure began to form.

With butadione in great excess, more oligomeric species can form, possibly even with the esterified boron centers depicted in Figure 6.8b. High concentrations of the diketone might overcome the steric repulsion from the five-membered ring. Taking into consideration the order-of-addition studies and these ^{11}B titration studies, the assembly formed readily in the presence of the metal-bound guanidinylmethyl terpyridine ligand. LRMS also confirmed formation of the assembly (SI). These promising results for the non-oligomerizing model spurred efforts to study oligomerization with a small molecule diboronic acid.

6.2.4 Proton NMR of Oligomerization with Thiophene-2,5-diyldiboronic Acid (6.13)

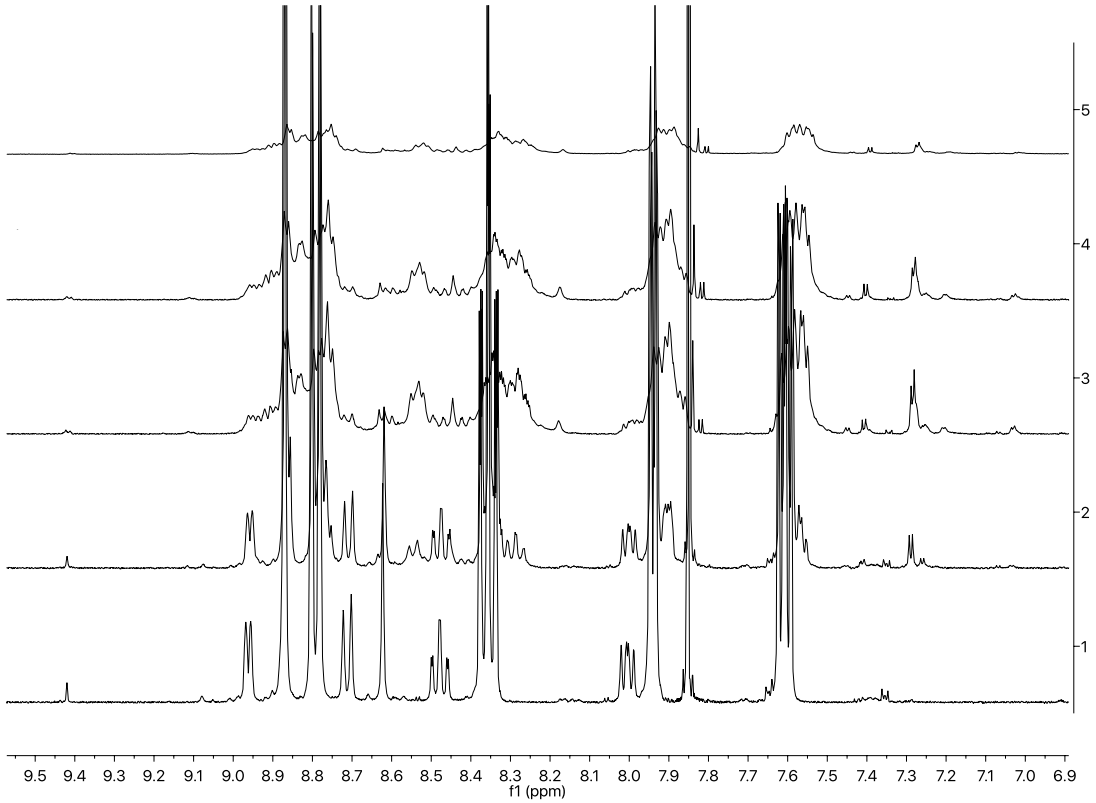


Figure 6.10: ^{1}H NMR Titration of Butadione into a Solution of 6.13 and Metal-Bound 6.1.

Oligomerization studies were done with compound 6.13 because of its small size and commercial availability. In addition, thiophenes display unique electronic properties

that could be helpful to our design.[25] Figure 6.10, a 1H NMR titration of Zn^{2+}-bound 6.1 and 6.13 with butadione demonstrates that as the equivalents of butadione increased, the aromatic thiophene protons' signal at 7.85 ppm decreased in intensity, and the decrease was attributed to boronic acid binding. Furthermore, a new singlet peak was not observed, perhaps because it disappeared under the other peaks. If discrete species had formed, a singlet should still have been observed. A small singlet did persist at 7.85 ppm, but it remains unclear if this was excess thiophene or partially bound 6.13. Nevertheless, these results suggest that the boronic acid electronic environment was changing.

6.2.5 DOSY NMR Studies

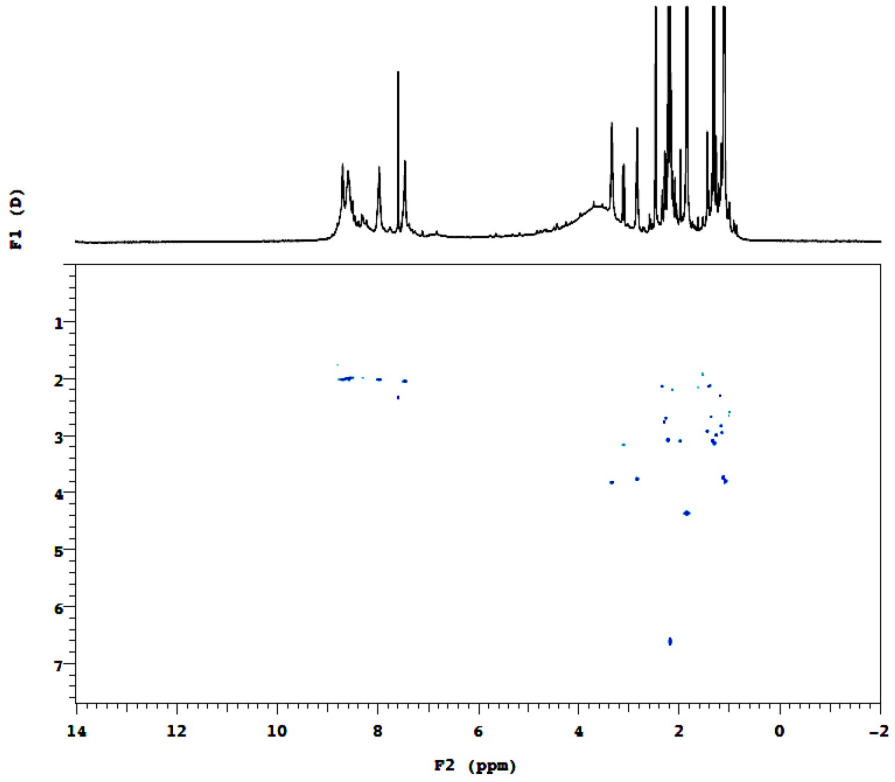


Figure 6.11: DOSY Experiment for a Mixture of Compounds 6.1, 6.13, and Butadione in DMSO.

To further probe if oligomerization occurred, DOSY experiments began with a control study in which metal-bound compound 6.7 and 6.13 were combined in deuterated

DMSO.. Five equivalents of DIPEA were added to ensure all pyridine rings were deprotonated and could complex with zinc. To start off, butadione was not added to the solution and the NMR experiment ran overnight at RT. As Figure 6.15 showed, aromatic peaks had a diffusion constant of approximately 2. When butadione was added, no change in the diffusion constant of the aromatic peak region was observed, indicating that an amine (not guanidine) could not promote oligomerization.

6.3 Conclusions

If chemists are to reliably form building blocks for unnatural oligomers, screen sequences of unnatural oligomers for numerous applications, and reliably reproduce the properties of specific sequences, a reconceptualization of oligomer design should take place. We argue that if we pursue simple building block design, emphasize straightforward reliable linking chemistries, and backbones that are readily modified, chemists can make oligomers that are easier to analyze for their intended purpose. In this work, we presented initial studies of guanidinium-terpyridine/metal/diketone/boronic acid oligomer geared towards achieving modular oligomer design. Future studies will include probing the oligomerization and possibly polymerizing this novel compound. More importantly, these studies were presented to emphasize the need for new approaches for making complex folding that strives to emulate structures known in nature.

6.4 Experimental

6.4.1 General Material

4'-chloro-2,2':6',2''-terpyridine was purchased from Acros and from Chem-Impex Inc.. 1,1'-Bis(diphenylphosphino)ferrocene was purchased from Combi-Blocks.. Tris(dibenzylideneacetone)palladium(0) was purchased from Oakwood Chemical. Pd/C, TFA, thiophene-2,5-diyldiboronic acid, deuterated ammonium acetate, $Zn(CN)_2$, and DIPEA were purchased from Sigma-Aldrich. Zinc triflate was purchased from Strem

Chemical. (2-(pyrrolidin-1-ylmethyl)phenyl)boronic acid and (2-((methylamino)methyl)phenyl)boronic acid were provided by the Anslyn group. Solvents were purchased from Fisher. D_2O, deuterated MeOD, and deuterated MeCN, and DMSO were purchased from Cambridge, Acros, and Sigma-Aldrich.

6.4.2 General Instrumentation

An Agilent NMR 400 was used for order of addition studies. A Varian DirectDrive 600 was used for boron and DOSY NMR experiments. An Agilent Technologies 6530 Accurate Mass QTofLC/MS was used for high-resolution mass spectrometry.). An Agilent Technologies 6130 Single Quadrupole LC/MS with Gemini C18 3.5 micron 2.1 50 mm was used for online separation (0.7 mL min^{-1} flow rate, 5–95% MeOH +0.1% formic acid in water over 12 min at room temperature) and low-resolution mass spectrometry. Wilmad precision 300 MHZ, OD 5mm, 7" length, class A glass, with a 0.38 mm thick wall, NMR tubes was used for DOSY experiments.

6.4.3 Synthesis of 1-([2,2':6',2''-terpyridin]-4'-ylmethyl)guanidine (6.1)

[2,2':6',2''-terpyridine]-4'-carbonitrile (6.6). 4'-chloro-2,2':6',2''-terpyridine (6.5) (15.1 mmole), $Zn(CN)_2$ (9.22 mmole), Zn dust (4.14 mmole), and bis(diphenylphosphino)ferrocene (dppf, 1.47 mmole) were dissolved in dimethylacetamide under argon. The solution was sparged with argon, followed by the introduction of Tris(dibenzylideneacetone)palladium(0) (4.13 mmole) ($Pd2(dba)_3$). The reaction was stirred at 180 °C for 2 hrs. Solvent was removed under vacuum with mild heating. Compound 6.6 was purified via flash column chromatography on neutral alumina gel, 10-15% EtOAc in hexanes. Yield: 10%. HRMS: found m/z 259.09760, calcd. 259.09780 $(M + H)^+$. ^1H NMR (400 MHz, $CDCl_3$) δ 8.65 (ddd, J = 4.8, 1.8, 0.9 Hz, 2H), 8.55 (s, 2H), 8.47 (dt, J = 7.9, 1.1 Hz, 2H), 7.83 (td, J = 7.7, 1.8 Hz, 2H), 7.34 (ddd, J = 7.6, 4.8, 1.2 Hz,

2H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 156.35, 153.88, 149.29, 137.01, 124.70, 122.18, 122.09, 122.03, 121.08, 116.85, 77.45, 77.13, 76.81.

[2,2':6',2''-terpyridin]-4'-ylmethanamine (6.7). Compound 6.6 (1.47 mmole) was dissolved in AcOH, 10% Pd/C (0.507 mmole) was added, and the reaction mixture was kept under H_2 balloon at RT for 12 hrs. Pd/C was then filtered off over hardened filter paper. Solvent was removed under vacuum with mild heating. No further purification was required. Yield: 90%. HRMS: found m/z 285.1110, calcd. 286.11400 $(M + Na)^+$. ^{1}H NMR (400 MHz, D_2O) δ 8.56 (dt, *J* = 4.8, 1.3 Hz, 2H), 8.25 (dt, *J* = 7.9, 1.1 Hz, 2H), 8.06 (s, 2H), 7.95 (td, *J* = 7.8, 1.8 Hz, 2H), 7.48 (ddd, *J* = 7.6, 5.0, 1.2 Hz, 2H), 4.30 (s, 2H).

Boc-1-([2,2':6',2''-terpyridin]-4'-ylmethyl)guanidine (6.9). Compound 6.7 (1.32 mmole) and *tert*-butyl (*Z*)-(((*tert*-butoxycarbonyl)amino)(1*H*-imidazol-1-yl)methylene)carbamate (1.45 mmole) (6.8) were dissolved in DCM, followed by addition of DIPEA (1.98 mmole). The reaction was stirred at RT for 1 hr, and then solvent was removed by rotary evaporation. The crude product was purified by flash column chromatography on neutral alumina gel, 5-20% EtOAc in hexanes. Yield: 50%. HRMS: found m/z 527.23820, calcd. 527.23770 $(M + Na)^+$. ^{1}H NMR (400 MHz, $CDCl_3$) δ 11.57 (s, 1H), 8.77 (t, *J* = 5.5 Hz, 1H), 8.68 (ddd, *J* = 4.8, 1.8, 0.9 Hz, 2H), 8.59 (dt, *J* = 8.1, 1.1 Hz, 2H), 8.39 (s, 2H), 7.84 (td, *J* = 7.7, 1.8 Hz, 2H), 7.31 (ddd, *J* = 7.5, 4.8, 1.2 Hz, 2H), 4.83 (d, *J* = 5.5 Hz, 2H), 1.47 (d, *J* = 7.2 Hz, 17H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 163.56, 156.54, 155.68, 155.61, 153.22, 148.92, 148.80, 137.25, 124.03, 121.57, 120.11, 83.43, 79.53, 44.23, 44.19, 28.35, 28.13, 14.28.

1-([2,2':6',2''-terpyridin]-4'-ylmethyl)guanidine (6.1). Compound 6.9 (1.75 mmol) was dissolved in a solution mixture of DCM and TFA (20% v/v) and stirred overnight at RT. Solvent was removed by rotary evaporation, and the resulting gum was triturated with Et_2O to produce a slightly pink solid. No further purification was performed.

Yield: Quantitative. ^{1}H NMR (400 MHz, D_2O) δ 8.52 (ddd, *J* = 5.0, 1.8, 0.9 Hz, 2H), 8.09 (dt, *J* = 8.0, 1.1 Hz, 2H), 7.89 (td, *J* = 7.8, 1.8 Hz, 2H), 7.84 (d, *J* = 0.9 Hz, 2H), 7.46 (ddd, *J* = 7.6, 5.0, 1.2 Hz, 2H), 4.56 – 4.51 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 154.36, 153.94, 149.88, 148.64, 138.85, 125.13, 121.73, 120.54, 119.57, 117.63, 114.73.

6.4.4 Order of Addition Proton NMR Studies

Proton NMR studies were performed using a deuterated solution with the following solvents and boron standards: D_2O/CD_3CN (3:1) (v/v) or 100mM $D_3CCO_2ND_4$ with $NaBF_4$ (1mg/mL).

For each iteration involving order of addition studies, 0.009 mmol of compound 6.1, 0.0045 mmol of zinc triflate, 0.009 mmol of compound 6.11, and 0.009-0.09 mmole of compound 6.2 were combined.

6.4.5 Boron NMR Titration Studies

For ^{11}B NMR titration studies, 0.2 mmol of compound 6.1, 0.1 mmol zinc triflate, 0.2 mmol of compound 6.12, as described in section 6.4.4, were mixed in 0.67 M D_7 ammonium acetate. An NMR spectrum was taken after 1, 2, 4, 6, and 20 equivalent addition of compound 6.2.

6.4.6 Proton DOSY NMR Experiments

For DOSY, a control study was performed on 0.015 mmol compound 6.7 dissolved in deuterated DMSO with 0.0075 mmol zinc triflate, and 0.0075 mmol of compound 6.13. The experiment ran overnight at RT using NMR tube specified in section 6.4.2, ensuring 3.8 cm of the solution were introduced.

The following day 10 equivalents of butadione were added and the solution remained at RT for 8 hrs. The NMR experiment was run overnight at RT.

6.5 Acknowledgements

We thank our following funding sources: the National Institute of Health (5DP1GM106408), the Defense Advanced Research Projects Agency (DARPA, N66001-14-2-4051), and the Welch Regents Chair (F-0046).

We also thank Anslyn undergraduate researcher Doug Saunders for synthesizing the products required for these studies. A special thanks to Rogelio Escamilla for his input and his synthetic skills that help push these initial studies forward.

6.7 Experimental Characterization

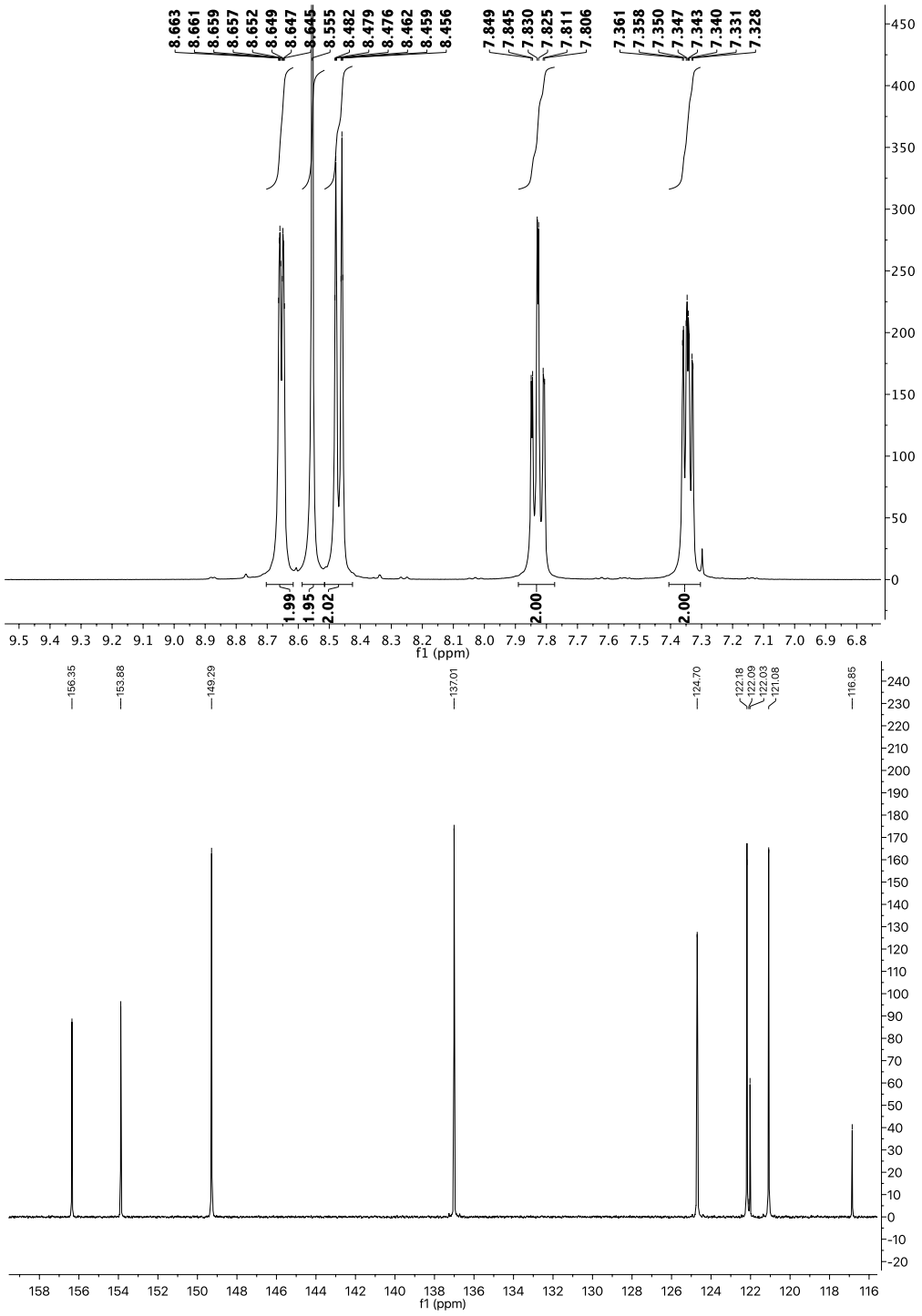
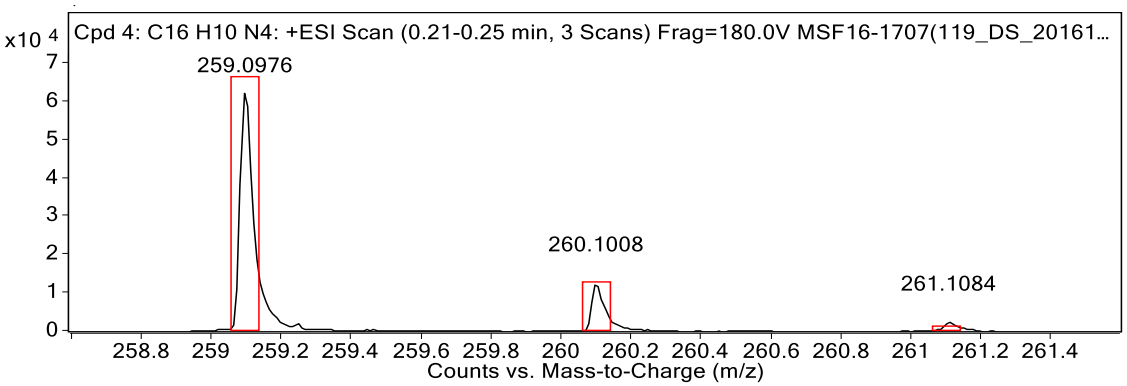


Figure 6.12: Proton (top) and carbon (bottom) for compound 6.6.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
259.09760	259.09780	1	64133.36	C16H10N4	(M+H)+	0.68
260.10080	260.10070	1	12783.25	C16H10N4	(M+H)+	-0.46
261.10840	261.10360	1	2534.61	C16H10N4	(M+H)+	-18.5
262.10380	262.10640	1	439.18	C16H10N4	(M+H)+	9.68

Figure 6.13: HRMS for compound for 6.6.

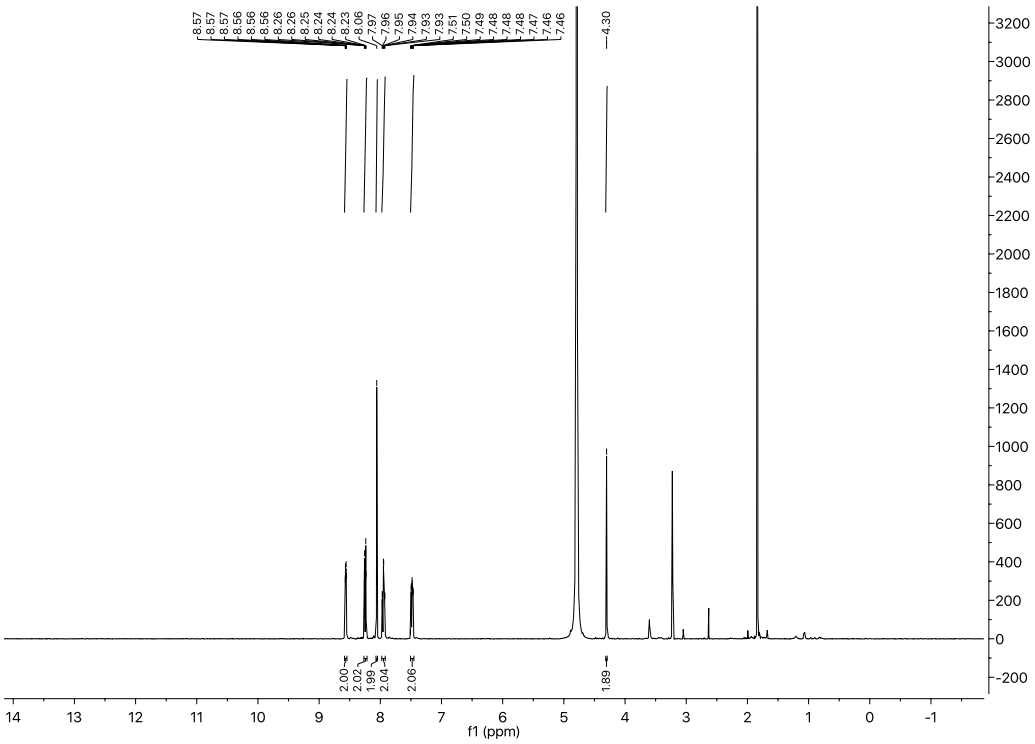
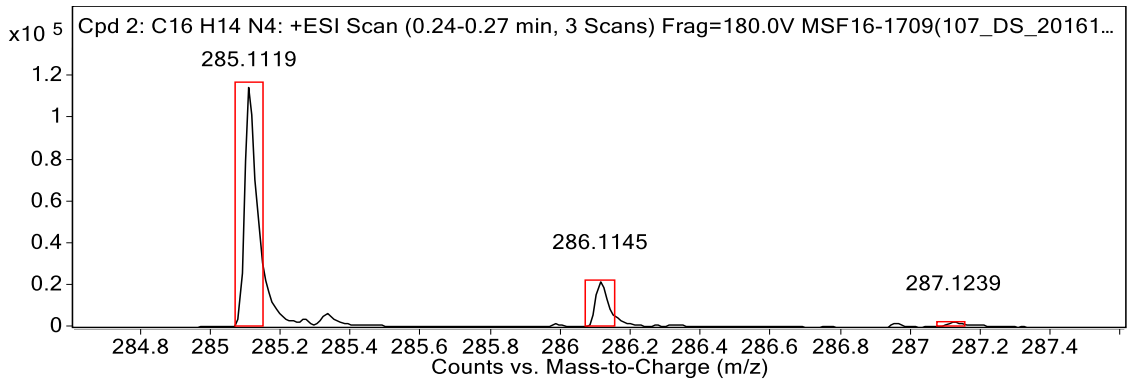


Figure 6.14: Proton NMR for 6.7.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
285.11190	285.11110	1	116079.98	C16H14N4	(M+Na)+	-2.76
286.11450	286.11400	1	22361.04	C16H14N4	(M+Na)+	-1.86
287.12390	287.11680	1	2616.77	C16H14N4	(M+Na)+	-24.6
659.28860			152518.16			

Figure 6.15: HRMS for compound for 6.7.

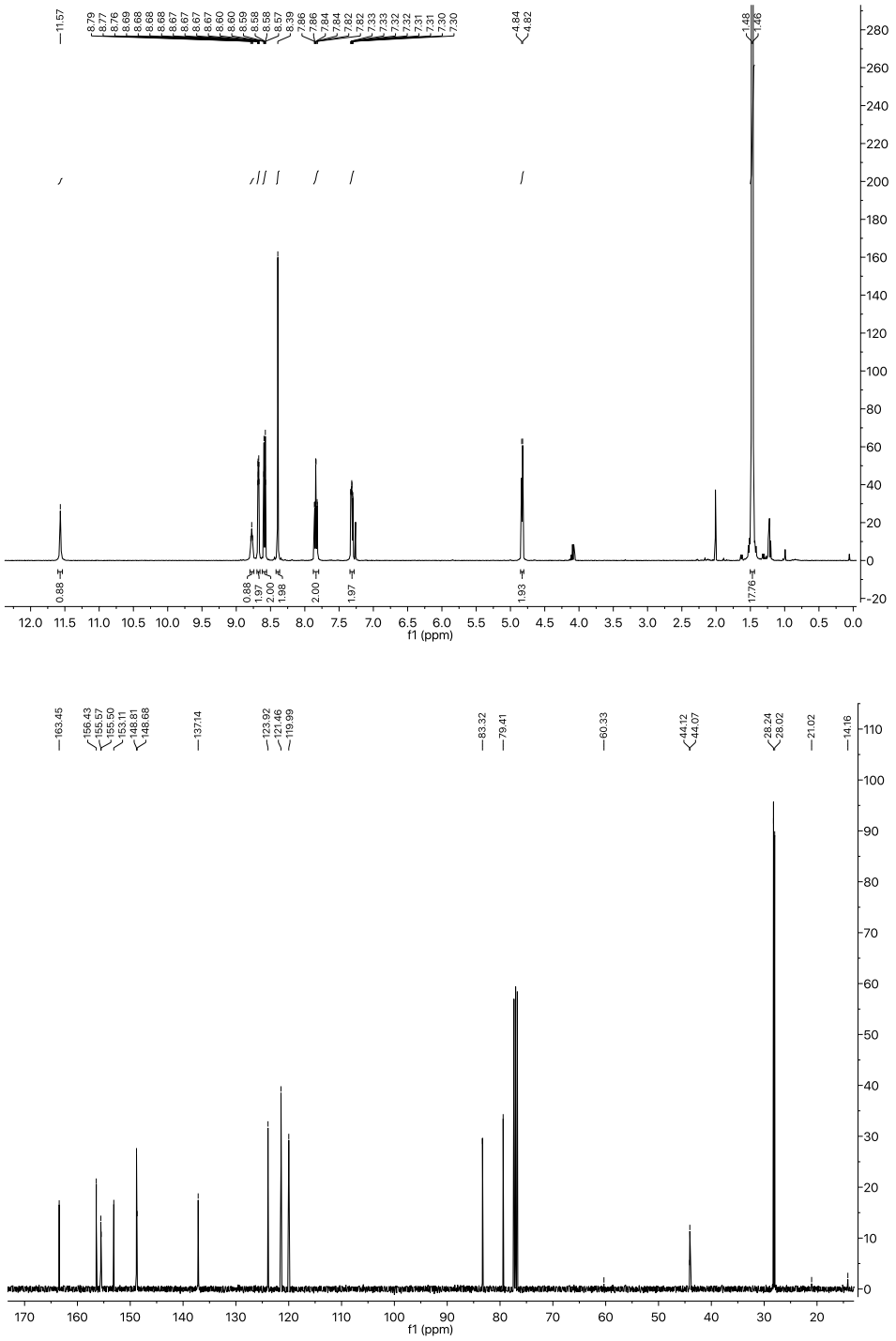
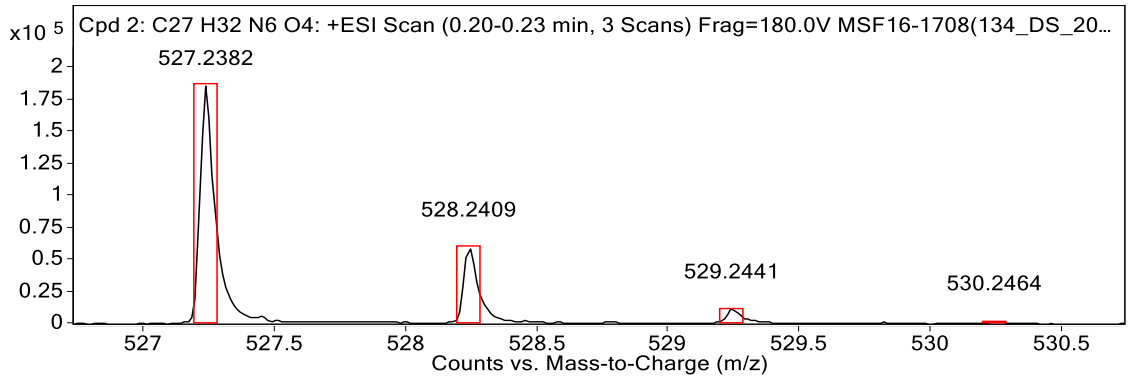


Figure 6.16: Proton (top) and carbon (bottom) for compound 6.9.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
505.25660			221405.38			
527.23820	527.23770	1	186138.77	C27H32N6O4	(M+Na)+	-0.97
528.24090	528.24070	1	59468.3	C27H32N6O4	(M+Na)+	-0.48
529.24410	529.24340	1	11178.94	C27H32N6O4	(M+Na)+	-1.33
530.24640	530.24600	1	1938.53	C27H32N6O4	(M+Na)+	-0.87

Figure 6.17: HRMS for compound for 6.9.

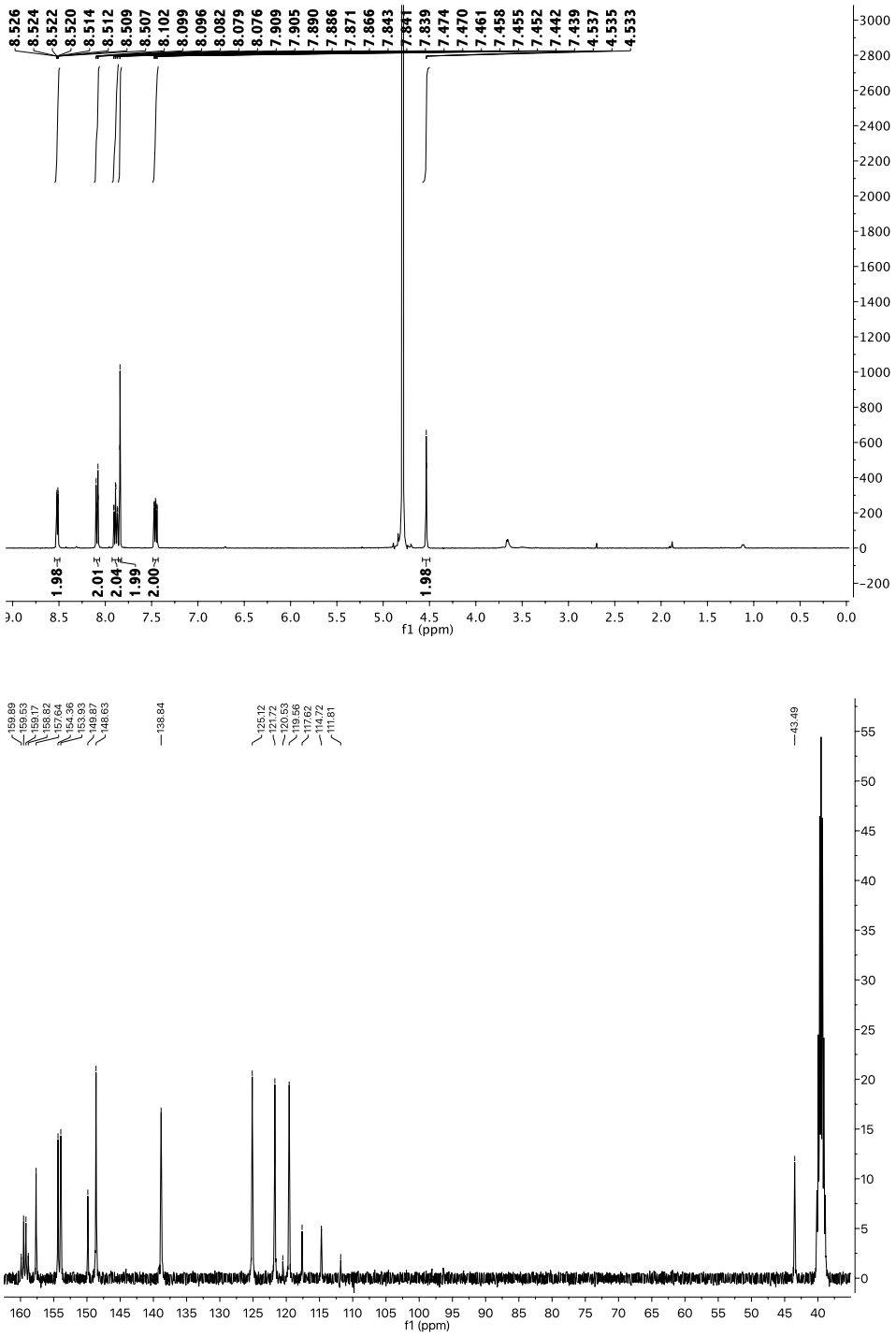
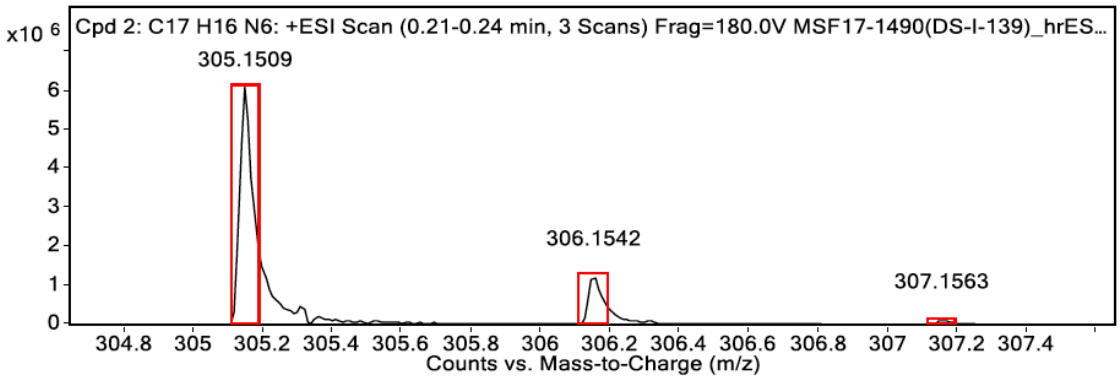


Figure 6.18: Proton (top) and carbon (bottom) for compound 6.1.



MS Spectrum Peak List

Obs. m/z	Calc. m/z	Charge	Abund	Formula	Ion/Isotope	Tgt Mass Error (ppm)
305.15090	305.15090	1	6124233.35	C17H16N6	(M+H)+	0.02
306.15420	306.15360	1	1232999.74	C17H16N6	(M+H)+	-1.75
307.15630	307.15630	1	127375.19	C17H16N6	(M+H)+	0.12
308.15840	308.15890	1	7587.01	C17H16N6	(M+H)+	1.64

Figure 6.19: HRMS for compound for 6.1.

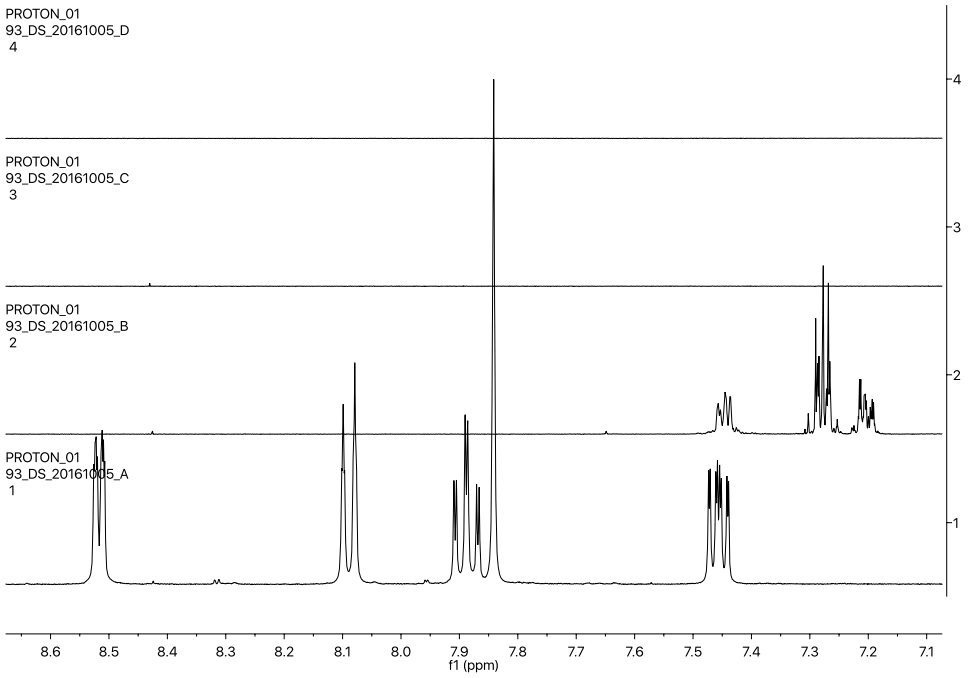


Figure 6.20: Proton NMR comparing compounds 6.1 (A), 6.11 (B), 6.2 (C), and Zn^{2+} (D).

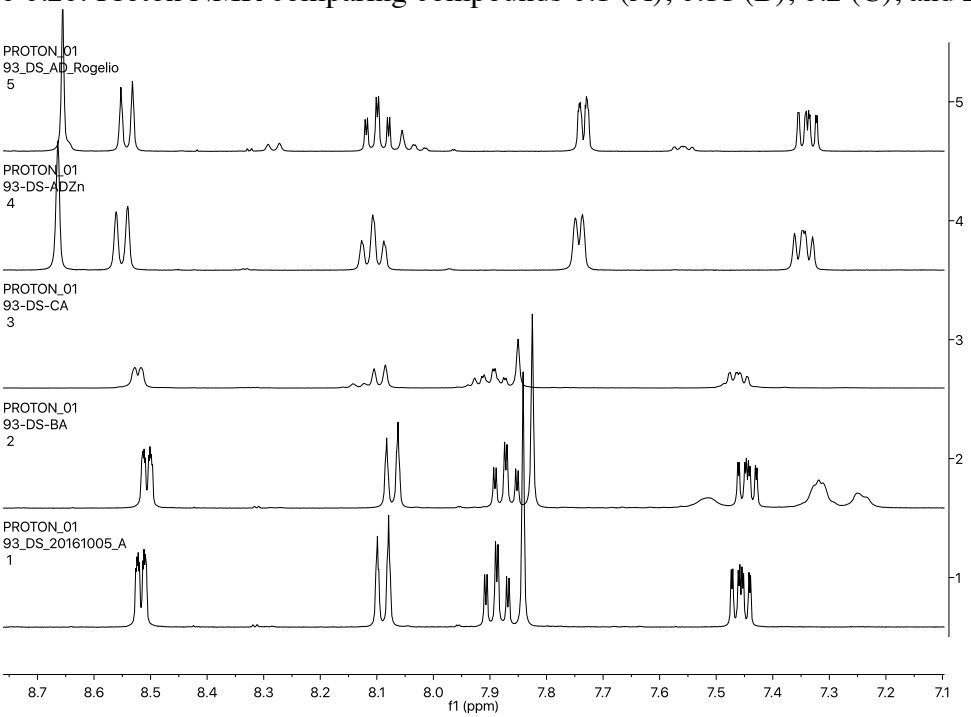


Figure 6.21: Proton NMR Comparing compounds 6.1 to B,A; C,A; D,A order of addition.

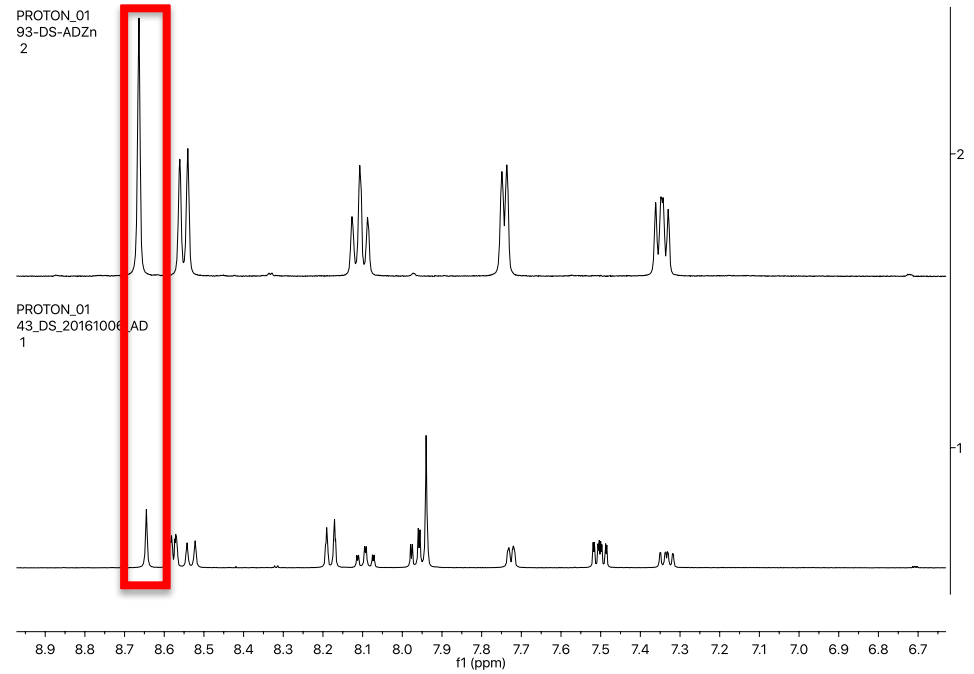


Figure 6.22: Proton NMR comparing aromatic region between fully/partially bound 6.1.

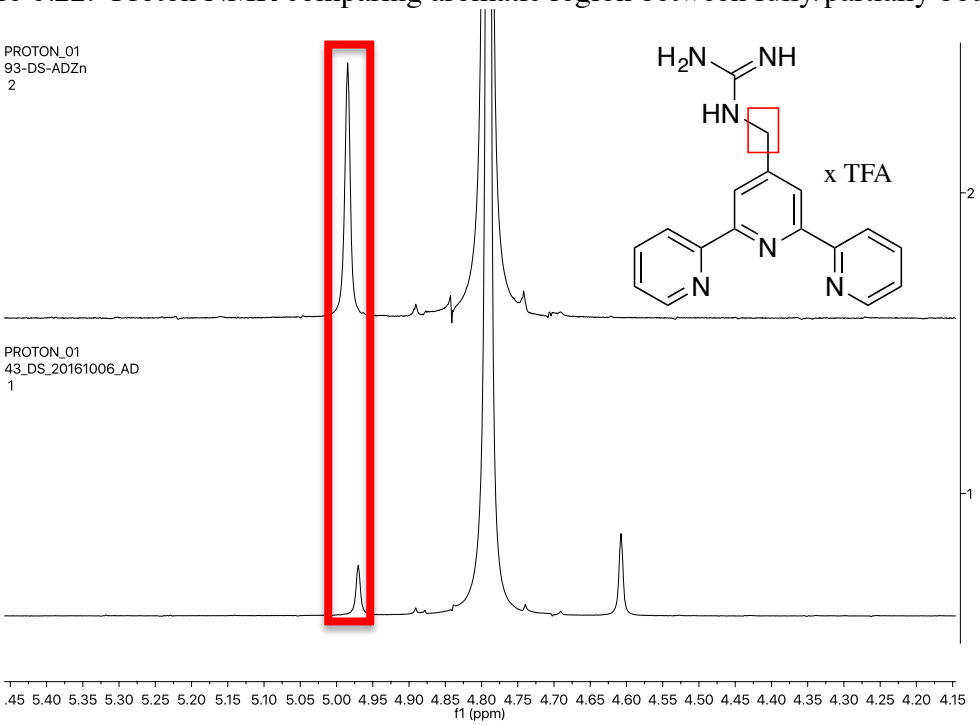


Figure 6.23: Proton NMR comparing guanidinyl methylene between fully/partially bound 6.1.

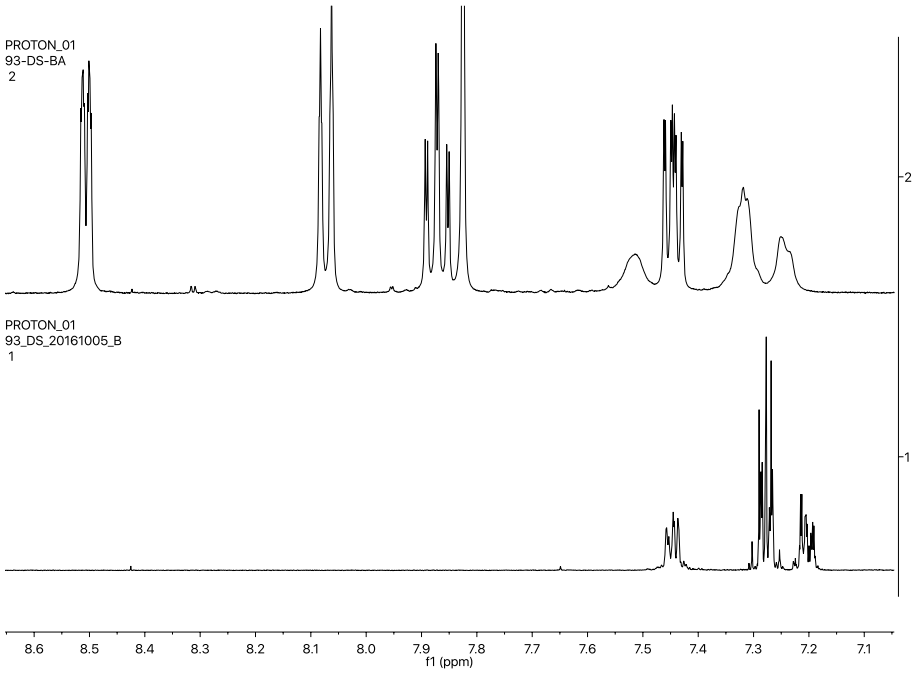


Figure 6.24: Proton NMR comparing compound 6.11 to a mixture of 6.11 and 6.1.

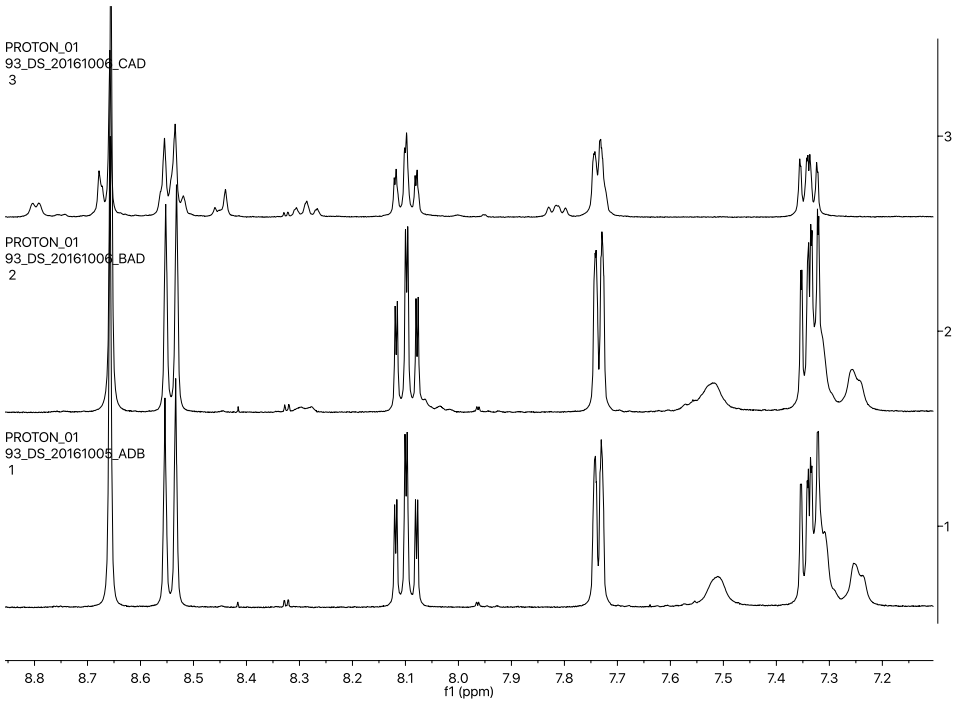


Figure 6.25: Proton NMR comparing orders of addition A, D, B; B, A, D; C, A, D for compounds 6.1, 6.11, and Zn^{2+}.

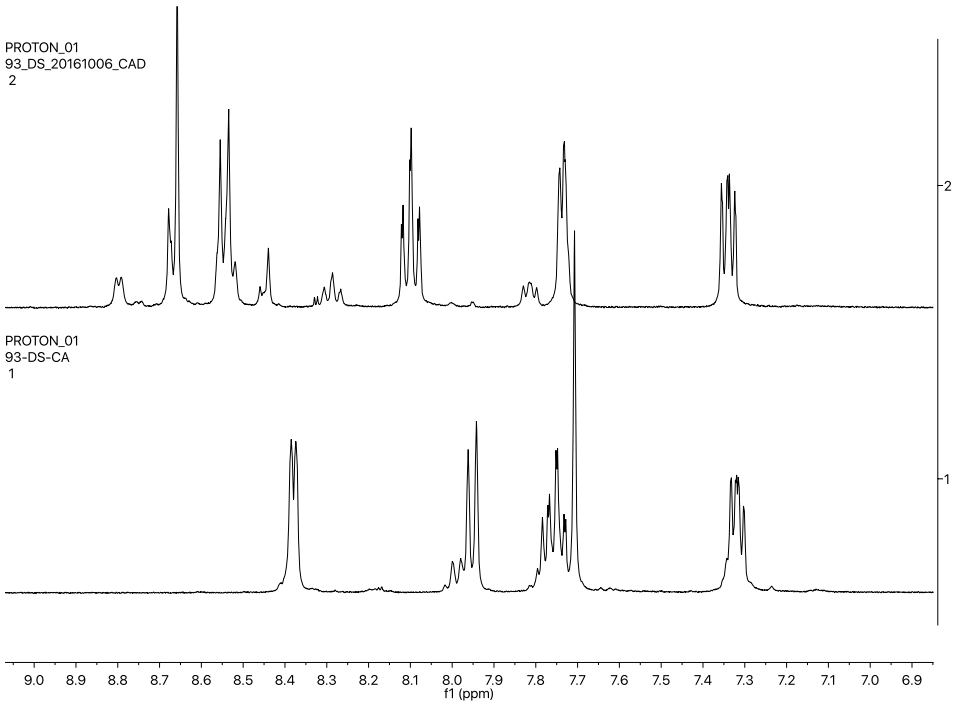


Figure 6.26: Proton NMR comparing orders of addition C, A; C, A, D for compounds 6.1, 6.2, and Zn^{2+}.

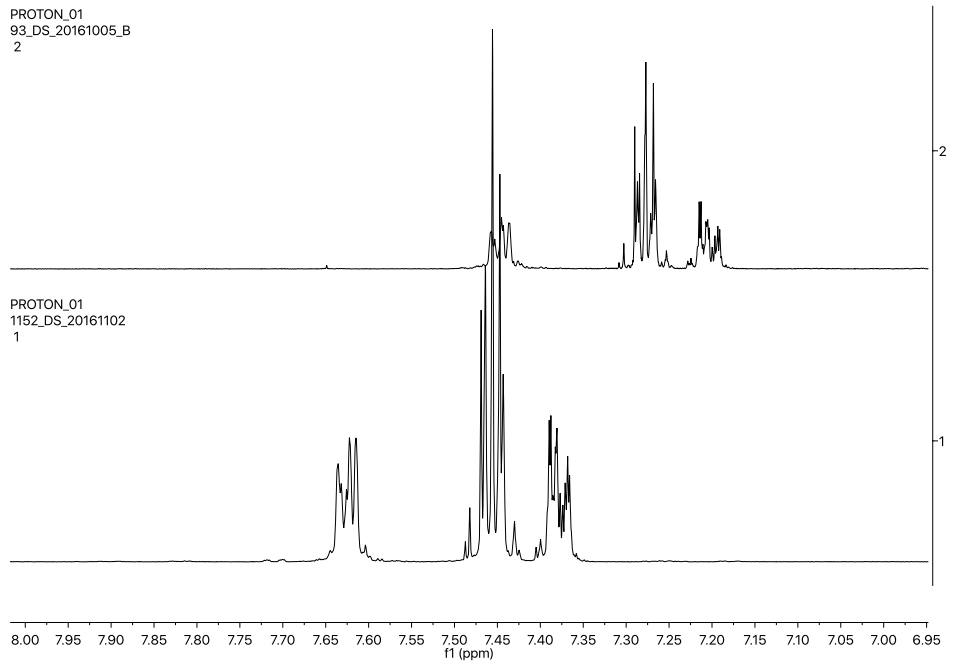


Figure 6.27: Proton NMR comparing compound 6.11 in the presence of Ac-Arg-OH. Changes in chemical shifts occurred due to differences solvents. Top spectra was more aqueous 1:1 (D2O:MeCN). Nevertheless, guanidinium did not broadened boronic aromatic region, suggesting interaction with terpyridine played an undetermined role influencing shift of boronic acid.

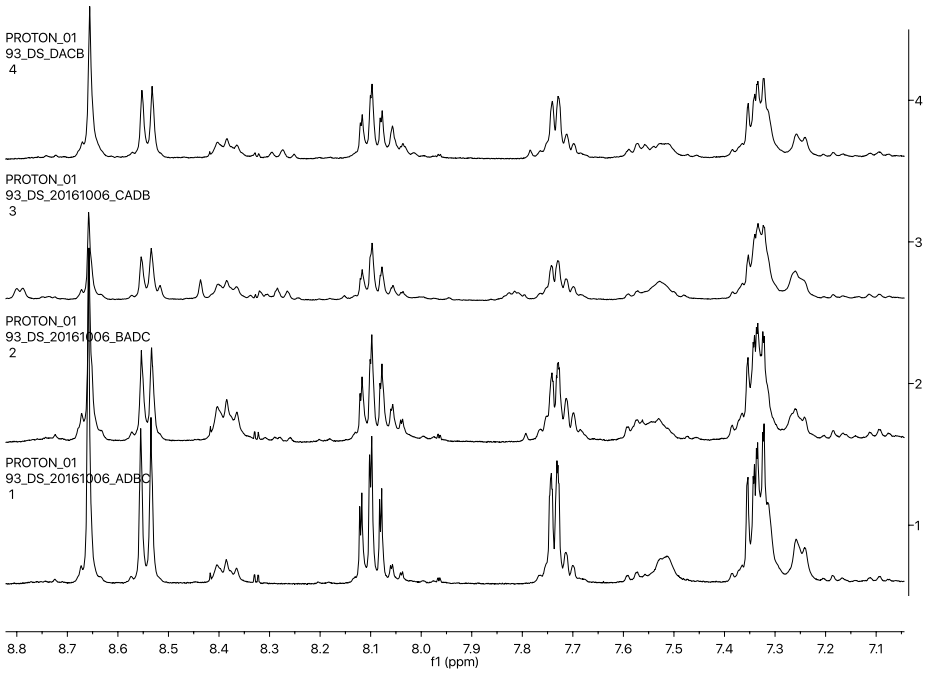


Figure 6.28: Proton NMR comparing all four orders of addition.

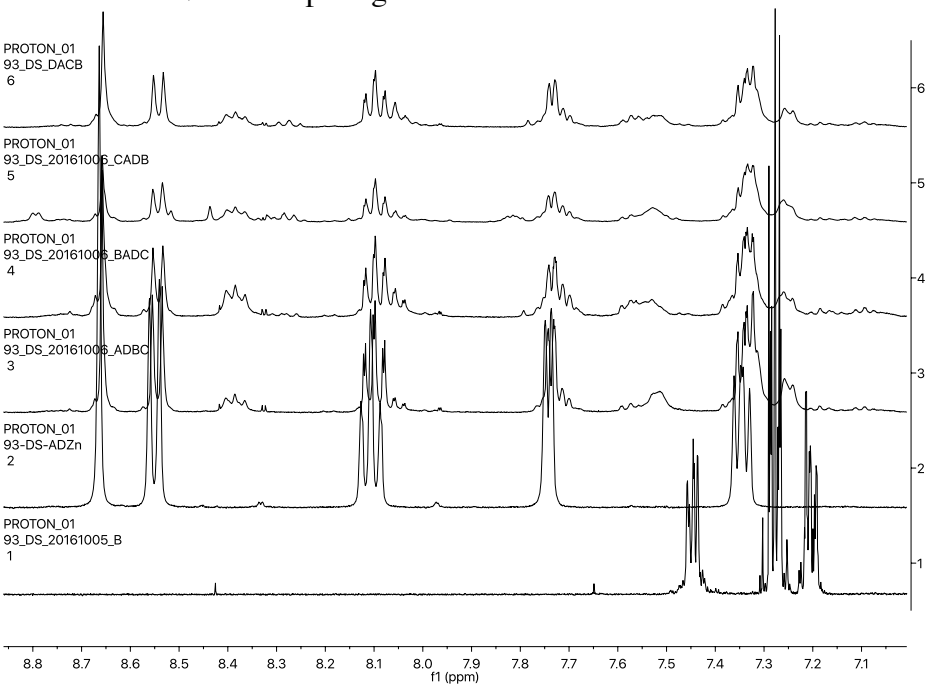


Figure 6.29: Proton NMR comparing compound 6.11 with all four orders of addition.

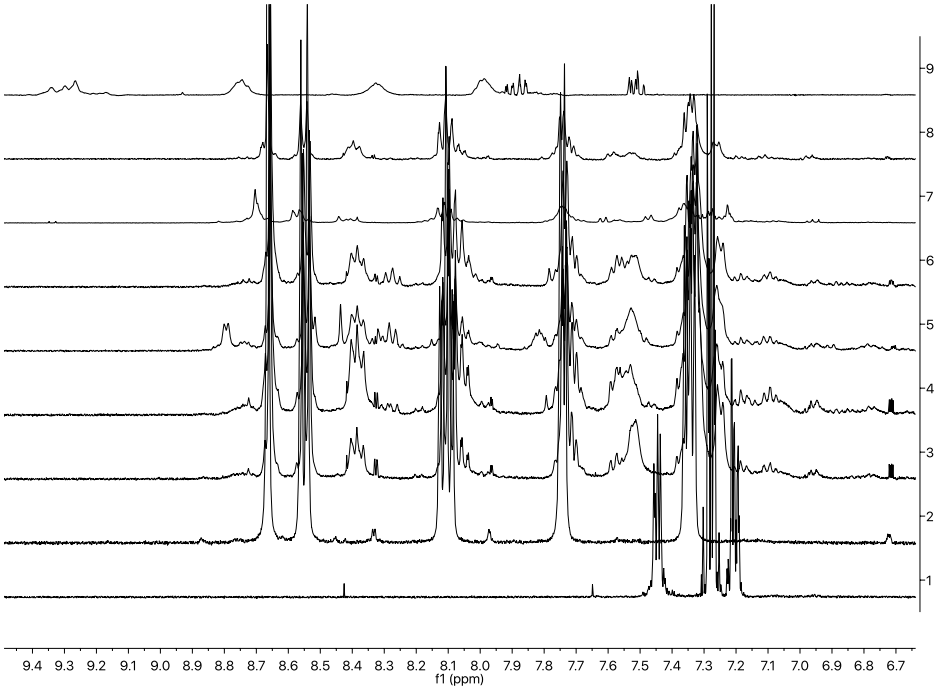


Figure 6.30: Proton NMR comparing compound 6.11 with all orders of addition and different conditions applied to increasing signal broadening described in section 6.2.2.

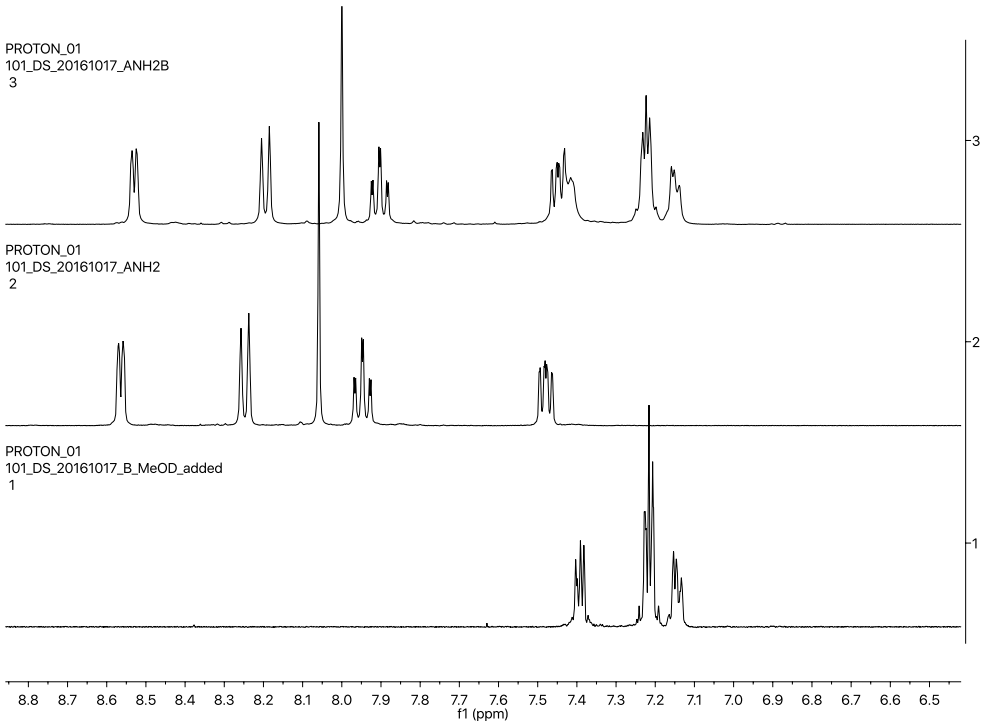


Figure 6.31: Proton NMR comparing compounds 6.9 and 6.11 individually and mixtures. Signal did not broadened when terpyridine or amine were present.

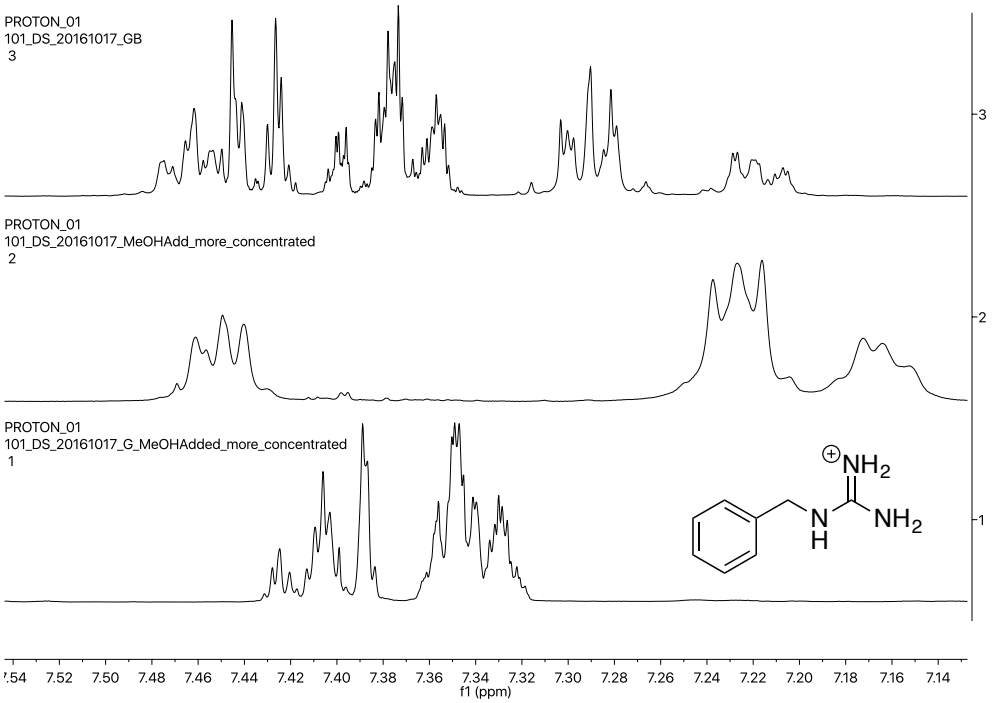


Figure 6.32: Proton NMR comparing benzyl guanidinium and compound 6.11. Due to solubility complications, this study did not definitively explain broadening of signal with compound 6.1.

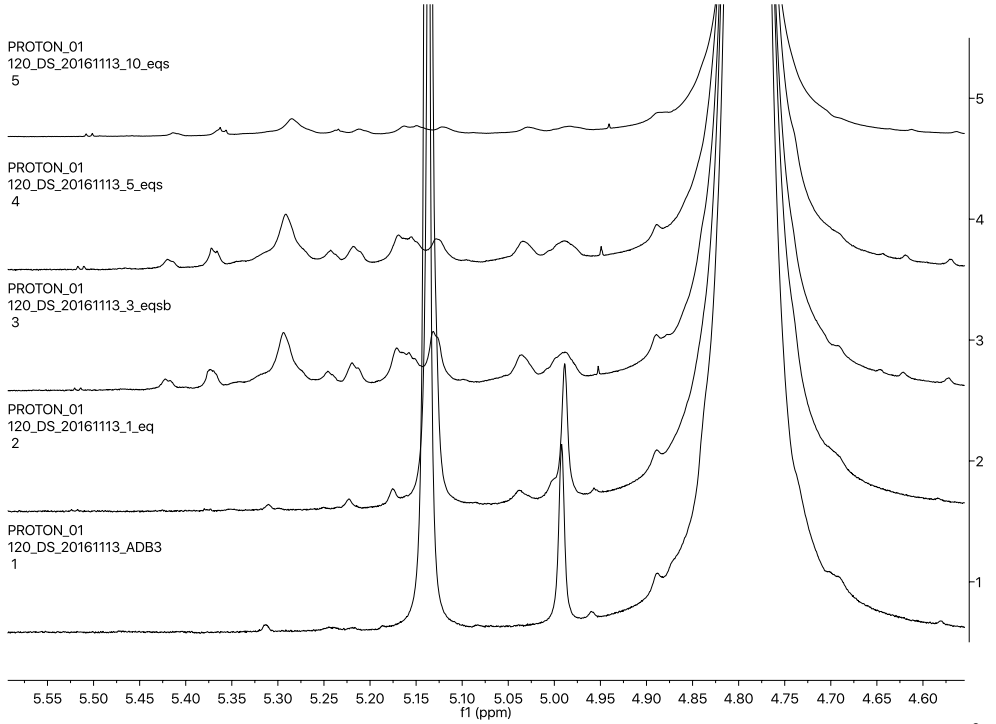


Figure 6.33: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.13, and Zn^{2+} for methylene group of 6.1.

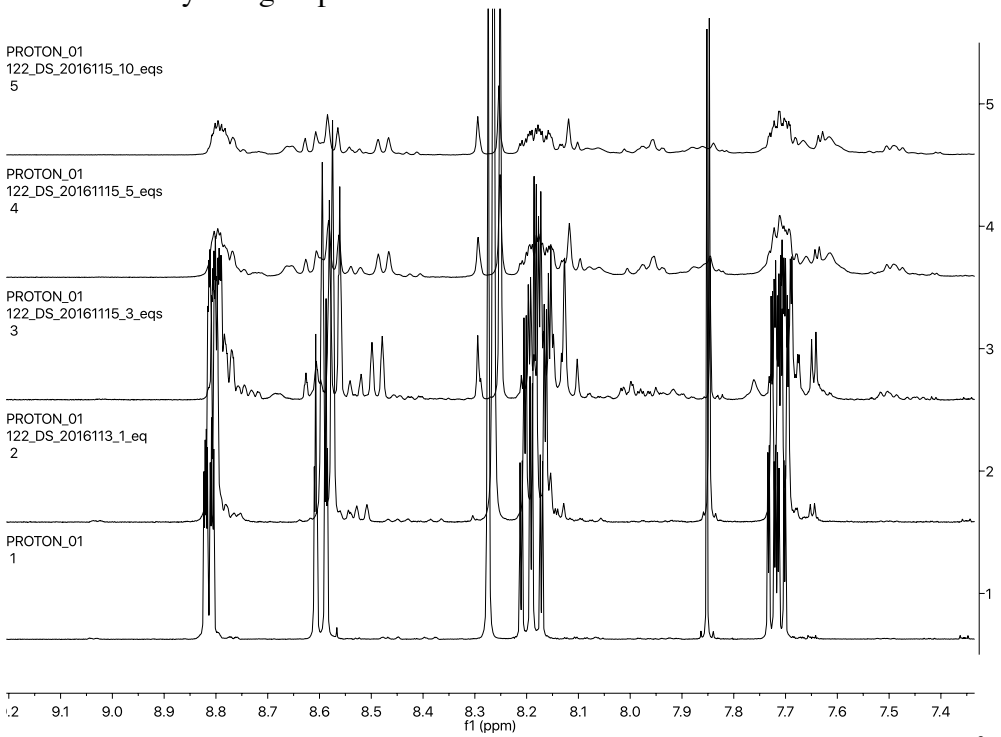


Figure 6.34: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.13, and Zn^{2+}. (Trial 2).

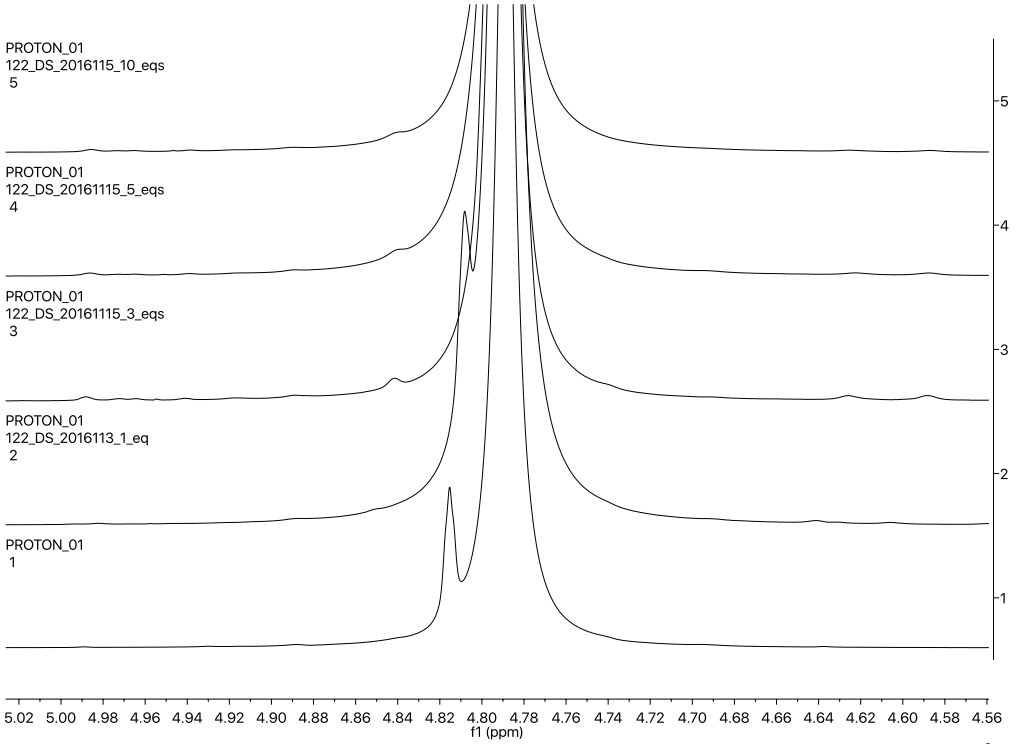


Figure 6.35: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.13, and Zn^{2+} for methylene group of 6.1.

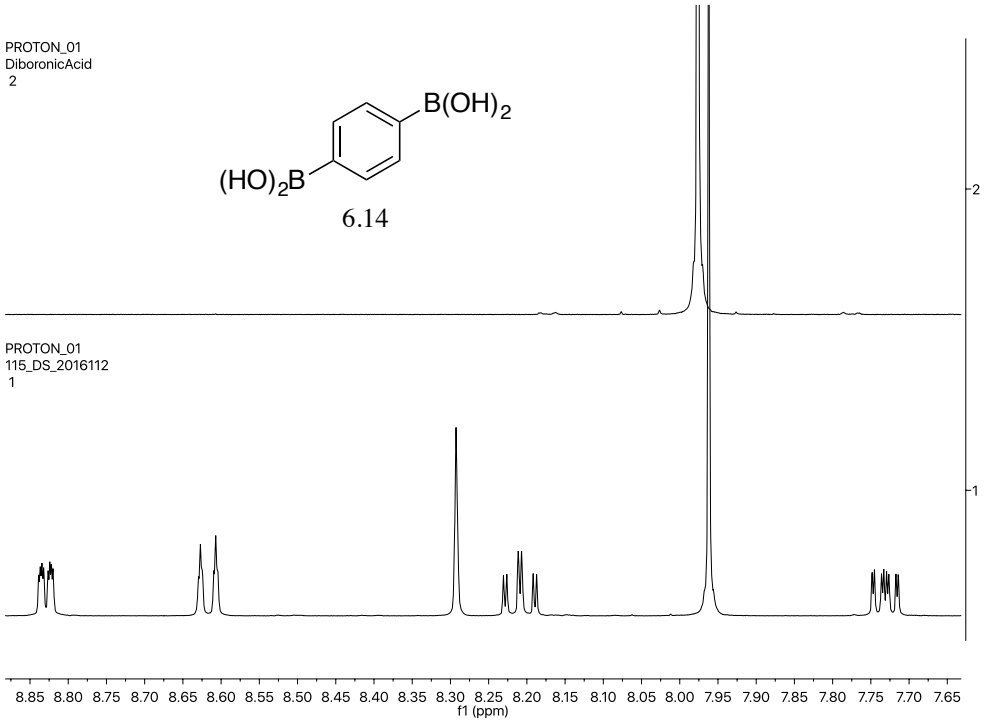


Figure 6.36: Comparing compound 6.1 and 6.14.

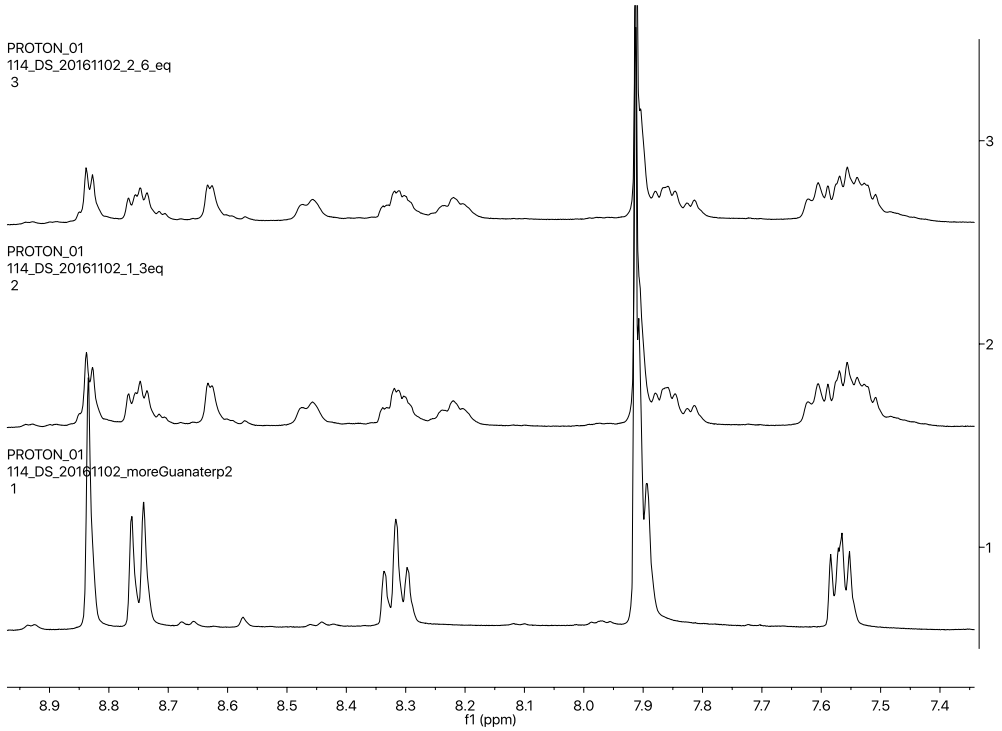


Figure 6.37: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.14, and Zn^{2+}.

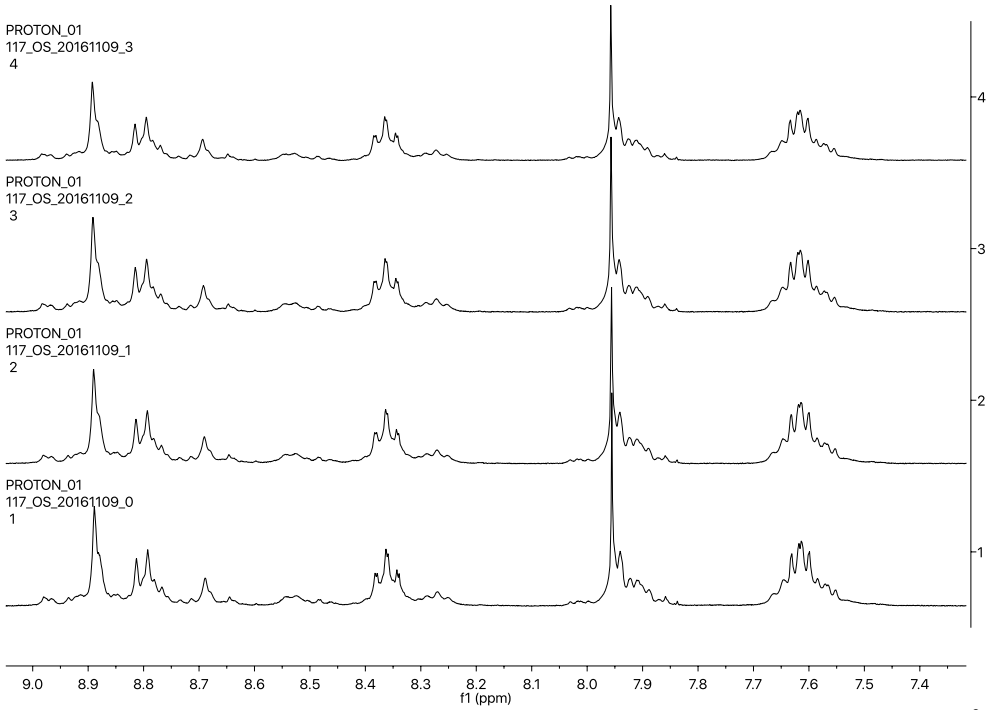


Figure 6.38: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.14, and Zn^{2+} (Trial 2).

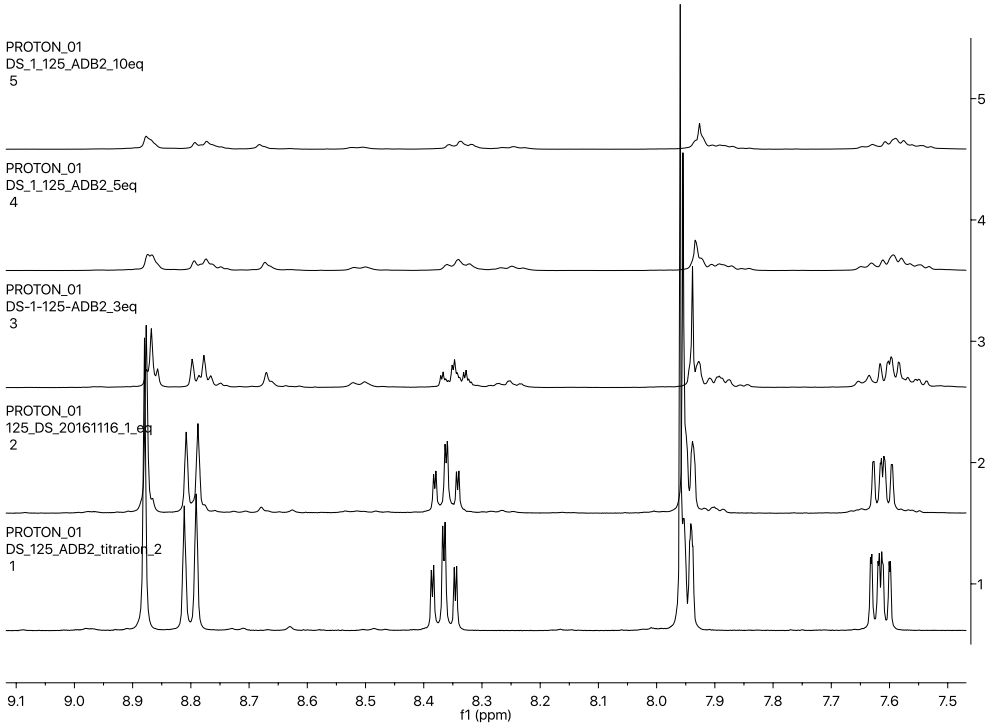


Figure 6.39: Proton NMR study of titrating 6.2 to a solution of 6.1, 6.14, and Zn^{2+} (Trial 3).

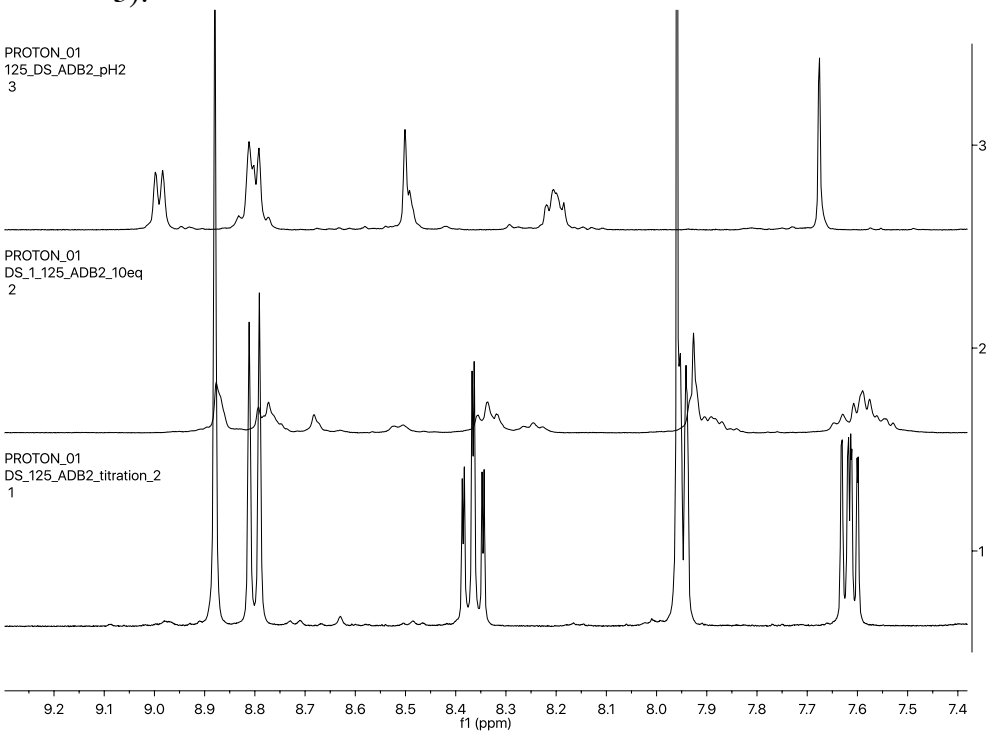


Figure 6.40: Proton NMR study comparing bound 6.14 and post acidification with deuterated TFA.

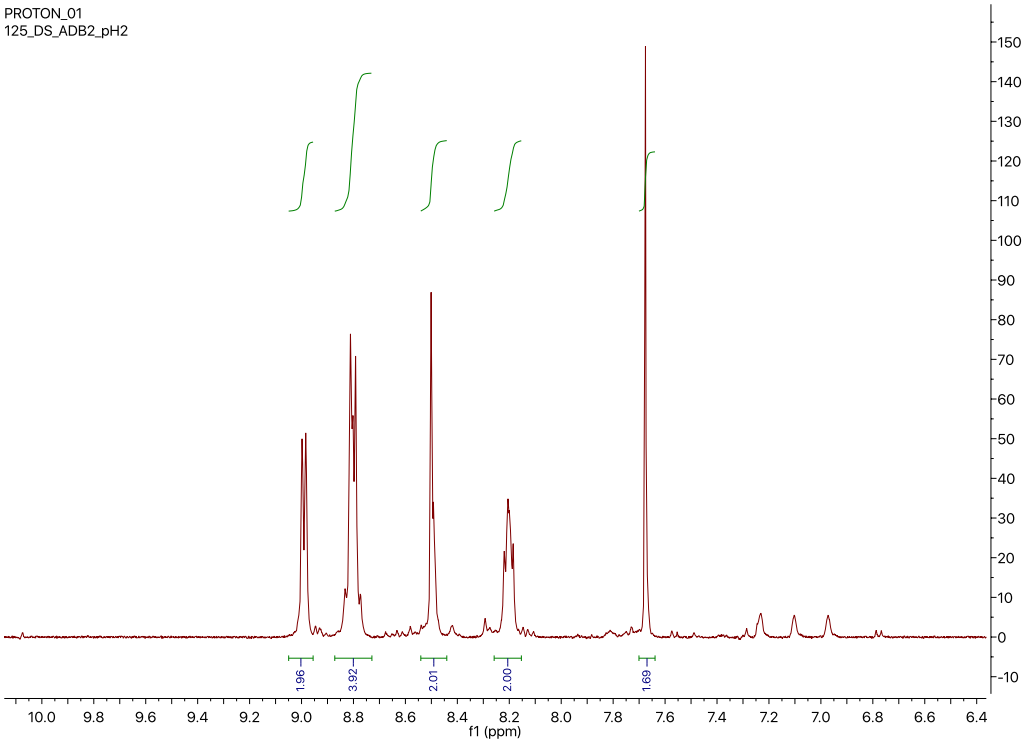


Figure 6.41: Proton NMR of 6.14 post acidification with deuterated TFA (refer to Figure 6.40). Stoichiometry of all species present previous to butadione addition persisted. However, it remained unclear if oligomer formed and so it DOSY studies were performed on compound 6.12 as described in section 6.2.4.

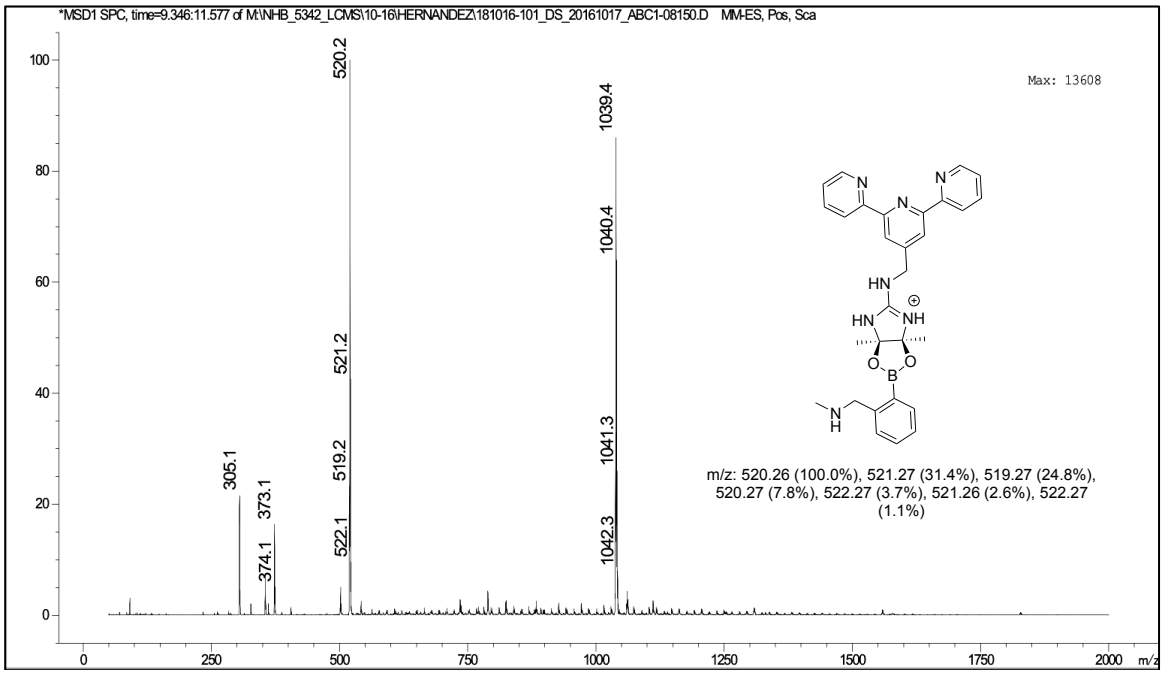


Figure 6.42: LRMS of assembly formed between compounds 6.1, 6.2, 6.11.

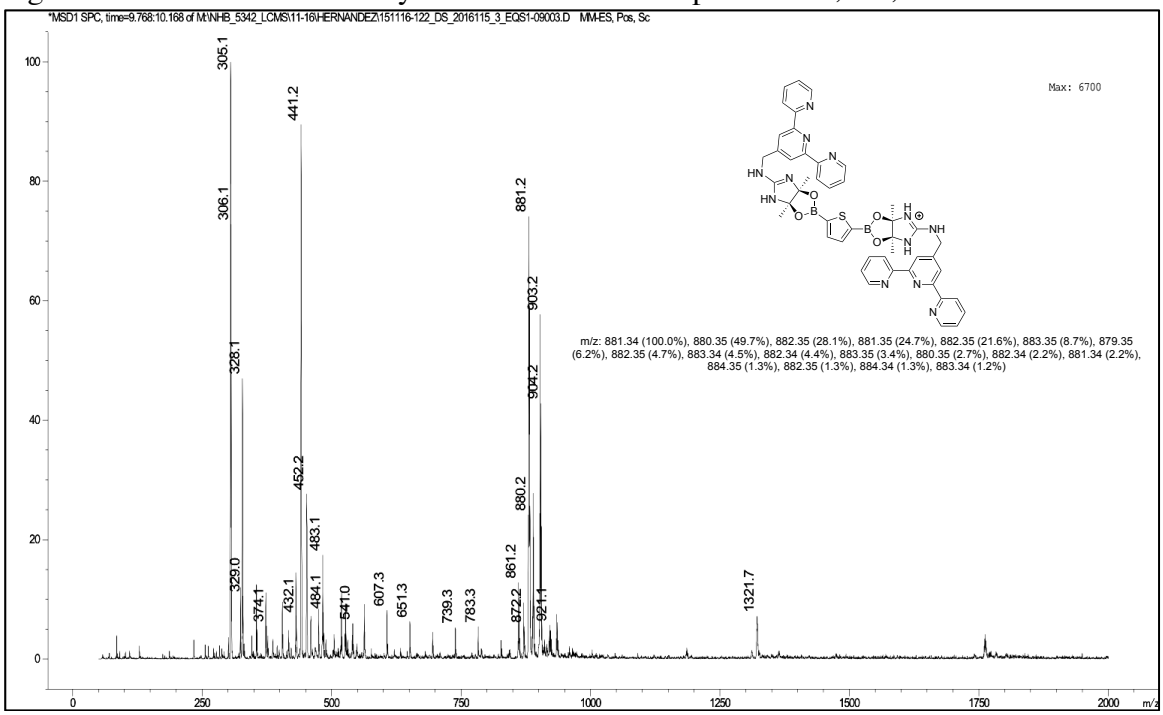


Figure 6.43: LRMS of assembly formed between compounds 6.1, 6.2, 6.13.

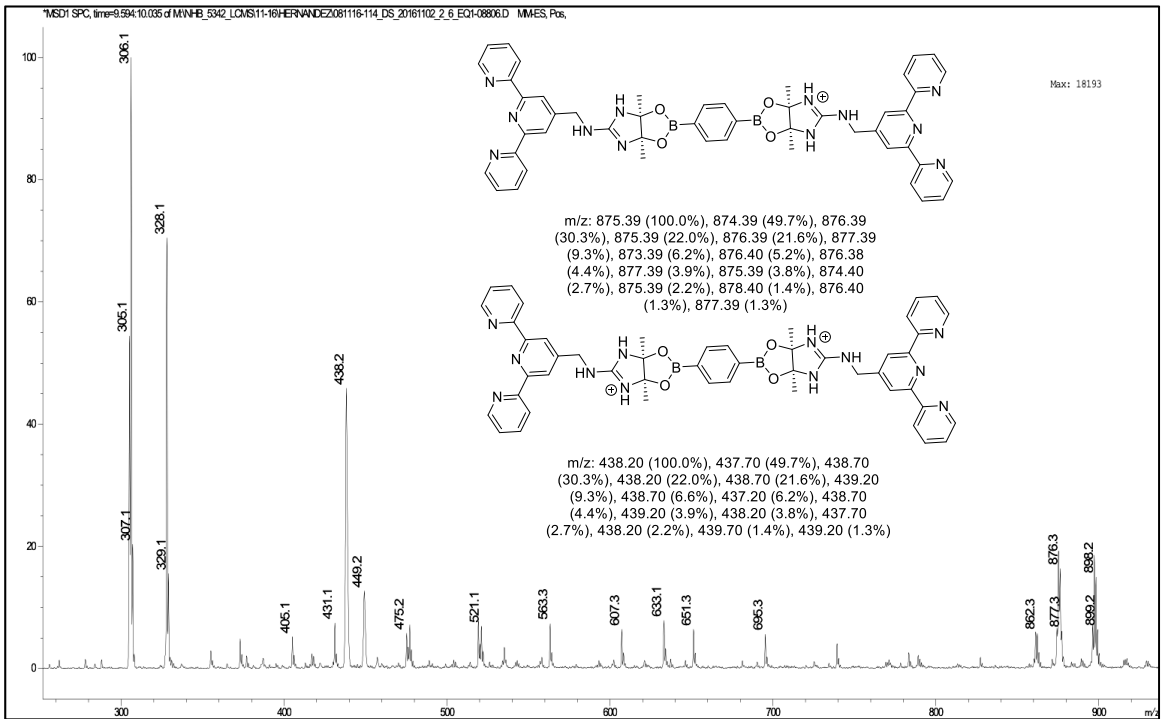


Figure 6.44: LRMS of assembly formed between compounds 6.1, 6.2, 6.14.

6.7 References

1. J. M. Palomo, *RSC Adv*., 2014, **4**, 32658-32672.
2. N. P. Chongsiriwatana, J. A. Patch, A. M. Czyzewski, M. T. Dohm, A. Ivankin, D. Gidalevitz, R. N. Zuckermann and A. E. Barron, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 2794-2799.
3. T. A. Martinek and F. Fulop, *Chem. Soc. Rev*., 2012, **41**, 687-702.
4. I. Huc and L. Cuccia, in *Foldamers*, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, DOI: 10.1002/9783527611478.ch1, pp. 1-33.
5. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore, *Chem. Rev*., 2001, **101**, 3893-4012.
6. W. S. Horne, M. D. Boersma, M. A. Windsor and S. H. Gellman, *Angew. Chem. Int. Ed*., 2008, **47**, 2853-2856.
7. P. Prabhakaran, G. Priya and G. J. Sanjayan, *Angew. Chem. Int. Ed.* , 2012, **51**, 4006-4008.
8. C. Nick Pace, J. M. Scholtz and G. R. Grimsley, *FEBS Lett*., 2014, **588**, 2177-2184.
9. E. Valeur and M. Bradley, *Chem. Soc. Rev*., 2009, **38**, 606-631.
10. V. Mäde, S. Els-Heindl and A. G. Beck-Sickinger, *Beilstein J. Org. Chem*., 2014, **10**, 1197-1212.
11. J. F. Riordan, *Biochemistry*, 1973, **12**, 3915-3923.

12. A. Leitner, S. Amon, A. Rizzi and W. Lindner, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 1321-1330.
13. A. Foettinger, A. Leitner and W. Lindner, *J. Chromatogr. A*, 2005, **1079**, 187-196.
14. K. L. Diehl, I. V. Kolesnichenko, S. A. Robotham, J. L. Bachman, Y. Zhong, J. S. Brodbelt and E. V. Anslyn, *Nat. Chem.*, 2016, **8**, 968-973.
15. L. Yang, X. Tan, Z. Wang and X. Zhang, *Chem. Rev.*, 2015, **115**, 7196-7239.
16. U. S. Schubert and C. Eschbaumer, *Angew. Chem. Int. Ed.*, 2002, **41**, 2892-2926.
17. H. Simon, G. Heubach, W. Bitterlich and H. Gleinig, *Ber. Dtsch. Chem. Ges*, 1965, **98**, 3692-3702.
18. C. Mousset, O. Provot, A. Hamze, J. Bignon, J.-D. Brion and M. Alami, *Tetrahedron*, 2008, **64**, 4287-4294.
19. F. Albericio and J. Tulla-Puche, in *The Power of Functional Resins in Organic Synthesis*, Wiley-VCH Verlag GmbH & Co. KGaA, 2009, DOI: 10.1002/9783527626175.ch1, pp. 1-14.
20. V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471-479.
21. A. Wilson, G. Gasparini and S. Matile, *Chem. Soc. Rev.*, 2014, **43**, 1948-1962.
22. H. M. Seifert, K. Ramirez Trejo and E. V. Anslyn, *J. Am. Chem. Soc.*, 2016, **138**, 10916-10924.
23. B. G. G. Lohmeijer and U. S. Schubert, *Macromol. Chem. Phys.*, 2003, **204**, 1072-1078.
24. R. Nishiyabu, Y. Kubo, T. D. James and J. S. Fossey, *Chem. Comm.*, 2011, **47**, 1106-1123.
25. F. Yakuphanoglu and B. F. Şenkal, *Polym. Eng. Sci.*, 2009, **49**, 722-726.

References

Chapter 1

1. T. Kimmerlin and D. Seebach, *J. Pep. Res.*, 2005, **65**, 229-260.
2. G. Jung and A. G. Beck-Sickinger, *Angew. Chem. Int. Ed.*, 1992, **31**, 367-383.
3. Y. Tsuda and Y. Okada, in *Amino Acids, Peptides and Proteins in Organic Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2010, DOI: 10.1002/9783527631803.ch6, pp. 201-251.
4. L. A. Carpino, S. Ghassemi, D. Ionescu, M. Ismail, D. Sadat-Aalaee, G. A. Truran, E. M. E. Mansour, G. A. Siwruk, J. S. Eynon and B. Morgan, *Org. Process Res. Dev.*, 2003, **7**, 28-37.
5. C. J. Forsyth, *Nat. Chem.*, 2010, **2**, 252-254.
6. R. A. Khan, *Chem. Biol. Drug. Des.*, 2016, **88**, 884-888.
7. G. Luca, M. Rossella De and C. Lucia, *Curr. Pharm. Design*, 2010, **16**, 3185-3203.
8. D. B. F. Johnson, J. K. Takimoto, J. Xu, L. Wang and T. P. Begley, in *Wiley Encyclopedia of Chemical Biology*, John Wiley & Sons, Inc., 2007, DOI: 10.1002/9780470048672.wecb585.
9. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore, *Chem. Rev.*, 2001, **101**, 3893-4012.
10. C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat. Chem. Biol.*, 2007, **3**, 252-262.
11. T. A. Martinek and F. Fulop, *Chem. Sov. Rev.*, 2012, **41**, 687-702.
12. M. T. Weinstock, J. N. Francis, J. S. Redman and M. S. Kay, *Pep. Sci.*, 2012, **98**, 431-442.
13. in *Foldamers*, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, DOI: 10.1002/9783527611478.fmatter, pp. I-XXII.
14. J. Wu, G. An, S. Lin, J. Xie, W. Zhou, H. Sun, Y. Pan and G. Li, *Chem. Comm.*, 2014, **50**, 1259-1261.
15. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149-2154.
16. P. H. H. Hermkens, H. C. J. Ottenheijm and D. Rees, *Tetrahedron*, 1996, **52**, 4527-4554.
17. F. Albericio and J. Tulla-Puche, in *The Power of Functional Resins in Organic Synthesis*, Wiley-VCH Verlag GmbH & Co. KGaA, 2009, DOI: 10.1002/9783527626175.ch1, pp. 1-14.
18. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606-631.
19. K. Fosgerau and T. Hoffmann, *Drug Discov. Today*, 2015, **20**, 122-128.
20. J. M. Palomo, *RSC Adv.*, 2014, **4**, 32658-32672.
21. R. Karstad, G. Isaksen, B.-O. Brandsdal, J. S. Svendsen and J. Svenson, *J. Med. Chem.*, 2010, **53**, 5558-5566.
22. A. Stevenazzi, M. Marchini, G. Sandrone, B. Vergani and M. Lattanzio, *Bioorg. Med. Chem. Lett.*, 2014, **24**, 5349-5356.
23. A. Ahmad Fuaad, F. Azmi, M. Skwarczynski and I. Toth, *Molecules*, 2013, **18**, 13148.

24. K. E. Schwieter and J. N. Johnston, *J. Am. Chem. Soc.*, 2016, **138**, 14160-14169.
25. D. A. Dougherty and E. B. Van Arnam, *Chem. Bio. Chem.*, 2014, **15**, 1710-1720.
26. C. Han and J. Wang, *Chem. Phys. Chem.*, 2012, **13**, 1522-1534.
27. B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem.*, 1996, **4**, 5-23.
28. H. Li, R. Aneja and I. Chaiken, *Molecules*, 2013, **18**, 9797.
29. D. Zamora-Olivares, T. S. Kaoud, J. Jose, A. Ellington, K. N. Dalby and E. V. Anslyn, *Angew. Chem. Int. Ed.*, 2014, **53**, 14064-14068.
30. S. H. Gellman, *Acc. Chem. Res.*, 1998, **31**, 173-180.
31. I. Saraogi and A. D. Hamilton, *Chem. Soc. Rev.*, 2009, **38**, 1726-1743.
32. W. S. Horne, L. M. Johnson, T. J. Ketas, P. J. Klasse, M. Lu, J. P. Moore and S. H. Gellman, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 14751-14756.
33. J. S. Laursen, J. Engel-Andreasen and C. A. Olsen, *Accts. Chem. Res.*, 2015, **48**, 2696-2704.

Chapter 2

26. Swaminathan, A. A. Boulgakov, E. M. Marcotte, *PLoS Comp. Bio.*, 2015, **11(2)**, 1-17.
27. Y. Yao, M. Docter, J. V. Ginkel, and C. Joo, Phys. Biol., 2015., **12,** 1-6.
28. Julka, and F. Regnier, *J. Proteome Res.*, 2004, **3**, 350-363.
29. S. L. Cockrill, K. L. Foster, J. Wildsmith, A. R. Goodrich, J. G. Dapron, T. C. Hassel, W.K. Kappel, and G. B. I. Scott, *Biotech.*, 2005, **38**, 301-304.
30. B. L. Frey, D. T. Ladror, S. B. Sondalle, C.J. Krusemark, A. L. Jue, J. J. Coon, and L. M. Smith, *J. Am. Mass Spectrom*, 2013, **24**, 1710-1721.
31. C.J . Krusemark, B. L. Frey, L. M. Smith, and P. J. Belshaw, *Gel-Free Proteomics, Methods in Molecular Biology*, **753** (Eds: Gevaert K, Vandekerckhove J) Humana Press, New York, 2011, pp.77-91.
32. H. R. Horton, and D. E. Koshland, *J. Am. Chem. Soc.* 1965, **87**, 1126-1132.
33. E. Scoffone, and A. Fontana, R. Rocchi, *Biochem*, 1968, **7**, 971-979.
34. H. Kuyama, M. Watanabe, C. Toda, E. Ando, K. Tanaka, and O. Nishimura, *Rapid Commun. Mass Spectrom.*, 2003, **17**, 1642-1650.
35. J. I. Macdonald, H. K. Munch, T. Moore, and M. B. Francis, *Nat. Chem. Bio.* 2015, **11**, 326-334.
36. M. Bantscheff, M. Bantscheff, G. Sweetman, and J. Rick, B. Kuster, *Anal. Bioanal. Chem.*, 2007, **389**, 1017-1031.
37. C. D. Spicer, and B. G. Davis, Nat. Comm. 2014, **5**, 1-14.
38. N. Krall, F. da Cruz, O. Boutureira, and G. J. L. Bernardes, *Nat. Chem.*, 2015, **8**, 1-11.
39. J. M. Chalker, and G. J. L. Bernardes, Y. A. Lin, B.G. Davis, *Chem. Asian J.* 2009, **4**, 630-640.
40. A. Isidro-Llobet, M. Alvarez, and F. Albericio, *Chem. Rev.,* 2009, **109**, 2455-2504.
41. B. J. Ko, and J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 1991-2000.
42. J. R.Moffet, and M. A.Namboodiri, *Immunol. Cell Biol.* 2003, **81**, 247-265.

43. E. Scoffone, A. Fontana, and R. Rocchi, *Biochem. Biophys. Res. Commun.*, 1966, **25**, 170-174.
44. V. Wittman, and S. Seeberger, *Angew. Chem. Int. Ed.*, 2000, **39**, 4348-4352.
45. J. Tulla-Puche, and F. Albericio, The (classic concept of) solid support. In *The power of functional resins in organic synthesis* (Eds: Tulla-Pucha J, Albericio F) Wiley, Weinheim, 2008, pp. 3-14.
46. C. R. Millington, R. Quarell, and G. Lowe, *Tett. Lett.*, 1998, **39**, 7201-7204.
47. C. Rosenbaum, and H. Waldmann, *Tett. Lett.* 2001, **42**, 5677-5680.
48. T. Keough, M. P. Lacey, and R. S. Yongquist, *Rapid Commun. Mass Spectrom.*, 2000, **14**, 2348-2356.
49. W. Tang, and M. L. Becker, *Chem. Soc. Rev.* 2014, **43**, 7013-7059.
50. L. Zervas, D. Borovas, and E. Gazis, *J. Am. Chem. Soc.* 1963, **85**: 3660-3666.

Chapter 3

1. M. Larance and A. I. Lamond, *Nat. Rev. Mol. Cell. Biol.*, 2015, **16**, 269-280.
2. Y. V. Karpievitch, A. D. Polpitiya, G. A. Anderson, R. D. Smith and A. R. Dabney, *Ann. Appl. Stat.*, 2010, DOI: 10.1214/10-AOAS341, 1797-1823.
3. N. Schupf, M. X. Tang, H. Fukuyama, J. Manly, H. Andrews, P. Mehta, J. Ravetch and R. Mayeux, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 14052-14057.
4. D. Theodorescu, E. Schiffer, H. W. Bauer, F. Douwes, F. Eichhorn, R. Polley, T. Schmidt, W. Schöfer, P. Zürbig, D. M. Good, J. J. Coon and H. Mischak, *Proteomics Clin. Appl.*, 2008, **2**, 556-570.
5. K. Petritis, L. J. Kangas, B. Yan, M. E. Monroe, E. F. Strittmatter, W.-J. Qian, J. N. Adkins, R. J. Moore, Y. Xu, M. S. Lipton, D. G. Camp and R. D. Smith, *Anal. Chem.*, 2006, **78**, 5026-5039.
6. J. S. Page, R. T. Kelly, K. Tang and R. D. Smith, *J. Am. Soc. Mass. Spectrom.*, 2007, **18**, 1582-1590.
7. W.-C. Yang, H. Mirzaei, X. Liu and F. E. Regnier, *Anal. Chem.*, 2006, **78**, 4702-4708.
8. Y. Gao and Y. Wang, *J. Am. Mass. Spec.*, 2007, **18**, 1973-1976.
9. J. Swaminathan, A. A. Boulgakov and E. M. Marcotte, *PLOS Computational Biology*, 2015, **11**, e1004080.
10. S. Goodwin, J. D. McPherson and W. R. McCombie, *Nat. Rev. Genet.*, 2016, **17**, 333-351.
11. J. M. Heather and B. Chain, *Genomics*, 2016, **107**, 1-8.
12. I. Braslavsky, B. Hebert, E. Kartalov and S. R. Quake, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 3960-3964.
13. J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. deWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K.

Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach and S. Turner, *Science*, 2009, **323**, 133-138.
14. S. Uemura, C. E. Aitken, J. Korlach, B. A. Flusberg, S. W. Turner and J. D. Puglisi, *Nature*, 2010, **464**, 1012-1017.
15. F. Ding, M. Manosas, M. M. Spiering, S. J. Benkovic, D. Bensimon, J.-F. Allemand and V. Croquette, *Nat. Meth,*, 2012, **9**, 367-372.
16. N. Bandeira, V. Pham, P. Pevzner, D. Arnott and J. R. Lill, *Nat. Biotech.*, 2008, **26**, 1336-1338.
17. Y. Yao, D. Margreet, G. Jetty van, R. Dick de and J. Chirlmin, *Phys. Biol.*, 2015, **12**, 055003.
18. T. Karstens and K. Kobs, *J. Phys. Chem.*, 1980, **84**, 1871-1872.
19. Y. A. Andreev, S. A. Kozlov, A. A. Vassilevski and E. V. Grishin, *Anal. Biochem.*, 2010, **407**, 144-146.
20. A. N. Zaykov, J. P. Mayer and R. D. DiMarchi, *Nat. Rev. Drug Discov.*, 2016, **15**, 425-439.
21. D. L. Swaney, C. D. Wenger and J. J. Coon, *J. Proteome Res.*, 2010, **9**, 1323-1329.
22. Y. Koide, Y. Urano, K. Hanaoka, T. Terai and T. Nagano, *J. Am. Chem. Soc.*, 2011, **133**, 5680-5682.
23. T. P. Hopp, K. S. Prickett, V. L. Price, R. T. Libby, C. J. March, D. Pat Cerretti, D. L. Urdal and P. J. Conlon, *Nat. Biotech.*, 1988, **6**, 1204-1210.

Chapter 4

1. S. Abdellaoui, B. C. Corgier, C. A. Mandon, B. Doumeche, C. A. Marquette and L. J. Blum, *Electroanal*, 2013, **25**.
2. R. Amin and S. A. Elfeky, *Spectrochim Acta A*, 2013, **108**.
3. S. Arimori, S. Ushiroda, L. M. Peter, A. T. A. Jenkins and T. D. James, *Chem. Commun.*, 2002, **20**.
4. J. B. Crumpton, W. Zhang and W. L. Santos, *Anal. Chem.*, 2011, **83**, 3548-3554.
5. W. W. Bachovchin, W. Y. L. Wong, S. Farr-Jones, A. B. Shenvi and C. A. Kettner, *Biochemistry (Mosc.)*, 1988, **27**, 7689-7697.
6. W. L. A. Brooks and B. S. Sumerlin, *Chem. Rev.*, 2016, **116**, 1375-1397.
7. J. N. Cambre and B. S. Sumerlin, *Polymer*, 2011, **52**, 4631-4643.
8. T. M. El Dine, J. Rouden and J. Blanchet, *Chem. Commun.*, 2015, **51**, 16084-16087.
9. T. D. James, K. R. A. S. Sandanayake and S. Shinkai, *Angew Chem Int Edit*, 1994, **33**.
10. S. Kotha, K. Lahiri and D. Kashinath, *Tetrahedron*, 2002, **58**.
11. K. Lacina and P. Skládal, *Electrochim Acta*, 2011, **56**.
12. K. Lacina, P. Skládal and T. D. James, *Chem. Cent. J.*, 2014, **8**, 1-17.
13. J. Langen, U. Fischer, M. Cavalar, C. Coetzee, P. Wegmann-Herr and H.-G. Schmarr, *Anal. Bioanal. Chem.*, 2016, **408**, 2425-2439.
14. J. Li, A. S. Grillo and M. D. Burke, *Accounts of Chemical Research*, 2015, **48**, 2297-2307.
15. M. Li, W. Zhu, F. Marken and T. D. James, *Chem. Commun.*, 2015, **51**, 14562-14573.

16. J. X. Qiao and P. Y. S. Lam, *Synthesis*, 2011, **6**.
17. L. Rocard, A. Berezin, F. De Leo and D. Bonifazi, *Angew. Chem. Int. Ed.*, 2015, **54**, 15739-15743.
18. G. Springsteen and B. Wang, *Chem. Commun.*, 2001, **17**.
19. X. Sun, W. Zhai, J. S. Fossey and T. D. James, *Chem. Commun.*, 2016, **52**, 3456-3469.
20. Q. Zhang, N. Tang, J. W. C. Brock, H. M. Mottaz, J. M. Ames, J. W. Baynes, R. D. Smith and T. O. Metz, *J. Proteome Res.*, 2007, **6**, 2323-2330.
21. W. Zhang, D. I. Bryson, J. B. Crumpton, J. Wynn and W. L. Santos, *Chem. Commun.*, 2013, **49**, 2436-2438.
22. X. Sun and T. D. James, *Chem. Rev.*, 2015, **115**, 8001-8037.
23. S.-H. Chung, T.-J. Lin, Q.-Y. Hu, C.-H. Tsai and P.-S. Pan, *Molecules*, 2013, **18**, 12346.
24. J. J. Deadman, S. Elgendy, C. A. Goodwin, D. Green, J. A. Baban, G. Patel, E. Skordalakes, N. Chino and G. Claeson, *J. Med. Chem.*, 1995, **38**, 1511-1522.
25. P. J. Duggan and D. A. Offermann, *Aust. J. Chem.*, 2007, **60**, 829-834.
26. N. Y. Edwards, T. W. Sager, J. T. McDevitt and E. V. Anslyn, *J. Am. Chem. Soc.*, 2007, **129**, 13575-13583.
27. G. A. Ellis, M. J. Palte and R. T. Raines, *J. Am. Chem. Soc.*, 2012, **134**, 3631-3634.
28. O. V. Gozhina, J. S. Svendsen and T. Lejon, *J. Pep. Sci.*, 2014, **20**, 20-24.
29. M. Kita, J. Yamamoto, T. Morisaki, C. Komiya, T. Inokuma, L. Miyamoto, K. Tsuchiya, A. Shigenaga and A. Otaka, *Tet. Lett.*, 2015, **56**, 4228-4231.
30. C.-H. Tsai, C.-H. Lin, C.-T. Hsieh, C.-C. Cai, T.-J. Lin, P.-Y. Liu, M.-H. Lin, M.-J. Wu, C.-C. Fu, Y.-C. Wu, F.-R. Chang and P.-S. Pan, *Res. Chem. Intermed.*, 2014, **40**, 2187-2198.
31. K. L. Bicker, J. Sun, J. J. Lavigne and P. R. Thompson, *ACS Comb. Sci.*, 2011, **13**, 232-243.
32. B. E. Collins and E. V. Anslyn, *Chem. European J.*, 2007, **13**, 4700-4708.
33. B. E. Collins, S. Sorey, A. E. Hargrove, S. H. Shabbir, V. M. Lynch and E. V. Anslyn, *J. Org. Chem.*, 2009, **74**, 4055-4060.
34. D. G. Hall, in *Boronic Acids*, Wiley-VCH Verlag GmbH & Co. KGaA, 2006, pp. 1-99.
35. J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem. .Asian J.*, 2009, **4**, 630-640.
36. N. Fomina, C. McFearin, M. Sermsakdi, O. Edigin and A. Almutairi, *J. Am. Chem. Soc.*, 2010, **132**, 9540-9542.
37. V. Castro, H. Rodríguez and F. Albericio, *ACS Comb. Sci.*, 2016, **18**, 1-14.

Chapter 5

1. M. C. Young, E. Liew, J. Ashby, K. E. McCoy and R. J. Hooley, *Chem. Comm.*, 2013, **49**, 6331-6333.
2. E. T. Hernandez, J. Swaminathan, E. M. Marcotte and E. V. Anslyn, *New J. Chem.*, 2017, **41**, 462-469.

3. E. T. Hernandez, I. V. Kolesnichenko, J. F. Reuther and E. V. Anslyn, *New J. Chem.*, 2017, **41**, 126-133.
4. C. Kaes, A. Katz and M. W. Hosseini, *Chem. Rev.*, 2000, **100**, 3553-3590.
5. T. Murashima, S. Tsukiyama, S. Fujii, K. Hayata, H. Sakai, T. Miyazawa and T. Yamada, *Org. Biomol. Chem.*, 2005, **3**, 4060-4064.
6. H. Ishida, M. Kyakuno and S. Oishi, *Pep. Sci.*, 2004, **76**, 69-82.
7. R. Zou, Q. Wang, J. Wu, J. Wu, C. Schmuck and H. Tian, *Chem. Soc. Rev.*, 2015, **44**, 5200-5219.
8. H. Ishida and Y. Inoue, *Pep. Sci.*, 2000, **55**, 469-478.
9. E. A. Ponomarenko, E. V. Poverennaya, E. V. Ilgisonis, M. A. Pyatnitskiy, A. T. Kopylov, V. G. Zgoda, A. V. Lisitsa and A. I. Archakov, *Int. J. Anal. Chem.*, 2016, **2016**, 6.
10. M. A. Zajac, *J. Org. Chem.*, 2008, **73**, 6899-6901.
11. L. Breydo and V. N. Uversky, *Metallomics*, 2011, **3**, 1163-1180.
12. H. M. Seifert, K. Ramirez Trejo and E. V. Anslyn, *J. Am. Chem. Soc.*, 2016, **138**, 10916-10924.
13. K. L. Diehl, J. L. Bachman, B. M. Chapin, R. Edupuganti, P. Rogelio Escamilla, A. M. Gade, E. T. Hernandez, H. H. Jo, A. M. Johnson, I. V. Kolesnichenko, J. Lim, C.-Y. Lin, M. K. Meadows, H. M. Seifert, D. Zamora-Olivares and E. V. Anslyn, in *Synthetic Receptors for Biomolecules: Design Principles and Applications*, The Royal Society of Chemistry, 2015, DOI: 10.1039/9781782622062-00039, pp. 39-85.
14. A. E. Sales, L. Breydo, T. S. Porto, A. L. F. Porto and V. N. Uversky, *RSC Adv.*, 2016, **6**, 42971-42983.
15. W. G. Skene and J.-M. P. Lehn, *Proc. Natl. Acad. Sci. U.S.A.*, 2004, **101**, 8270-8275.
16. R. J. Wojtecki, M. A. Meador and S. J. Rowan, *Nat. Mater.*, 2011, **10**, 14-27.
17. T. Ganguly, B. B. Kasten, D.-K. Bucar, L. R. MacGillivray, C. E. Berkman and P. D. Benny, *Chem. Comm.*, 2011, **47**, 12846-12848.
18. C.-H. Lu, Y.-F. Lin, J.-J. Lin and C.-S. Yu, *PLOS ONE*, 2012, **7**, e39252.
19. M. M. Yamashita, L. Wesson, G. Eisenman and D. Eisenberg, *Proceedings of the National Academy of Sciences*, 1990, **87**, 5648-5652.
20. J. M. Dragna, G. Pescitelli, L. Tran, V. M. Lynch, E. V. Anslyn and L. Di Bari, *J. Am. Chem. Soc.*, 2012, **134**, 4398-4407.
21. U. McDonnell, M. R. Hicks, M. J. Hannon and A. Rodger, *J. Inorg. Biochem.*, 2008, **102**, 2052-2059.
22. C. J. White and A. K. Yudin, *Nat. Chem.*, 2011, **3**, 509-524.
23. T. Gruendling, S. Weidner, J. Falkenhagen and C. Barner-Kowollik, *Poly. Chem.*, 2010, **1**, 599-617.
24. J. H. van Esch, M. A. M. Hoffmann and R. J. M. Nolte, *J. Org. Chem.*, 1995, **60**, 1599-1610.

Chapter 6

1. J. M. Palomo, *RSC Adv.*, 2014, **4**, 32658-32672.

2. N. P. Chongsiriwatana, J. A. Patch, A. M. Czyzewski, M. T. Dohm, A. Ivankin, D. Gidalevitz, R. N. Zuckermann and A. E. Barron, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 2794-2799.
3. T. A. Martinek and F. Fulop, *Chem. Soc. Rev.*, 2012, **41**, 687-702.
4. I. Huc and L. Cuccia, in *Foldamers*, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, DOI: 10.1002/9783527611478.ch1, pp. 1-33.
5. D. J. Hill, M. J. Mio, R. B. Prince, T. S. Hughes and J. S. Moore, *Chem. Rev.*, 2001, **101**, 3893-4012.
6. W. S. Horne, M. D. Boersma, M. A. Windsor and S. H. Gellman, *Angew. Chem. Int. Ed.*, 2008, **47**, 2853-2856.
7. P. Prabhakaran, G. Priya and G. J. Sanjayan, *Angew. Chem. Int. Ed.* , 2012, **51**, 4006-4008.
8. C. Nick Pace, J. M. Scholtz and G. R. Grimsley, *FEBS Lett.*, 2014, **588**, 2177-2184.
9. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606-631.
10. V. Mäde, S. Els-Heindl and A. G. Beck-Sickinger, *Beilstein J. Org. Chem.*, 2014, **10**, 1197-1212.
11. J. F. Riordan, *Biochemistry*, 1973, **12**, 3915-3923.
12. A. Leitner, S. Amon, A. Rizzi and W. Lindner, *Rapid Commun. Mass Spectrom.*, 2007, **21**, 1321-1330.
13. A. Foettinger, A. Leitner and W. Lindner, *J. Chromatogr. A*, 2005, **1079**, 187-196.
14. K. L. Diehl, I. V. Kolesnichenko, S. A. Robotham, J. L. Bachman, Y. Zhong, J. S. Brodbelt and E. V. Anslyn, *Nat. Chem.*, 2016, **8**, 968-973.
15. L. Yang, X. Tan, Z. Wang and X. Zhang, *Chem. Rev.*, 2015, **115**, 7196-7239.
16. U. S. Schubert and C. Eschbaumer, *Angew. Chem. Int. Ed.*, 2002, **41**, 2892-2926.
17. H. Simon, G. Heubach, W. Bitterlich and H. Gleinig, *Ber. Dtsch. Chem. Ges*, 1965, **98**, 3692-3702.
18. C. Mousset, O. Provot, A. Hamze, J. Bignon, J.-D. Brion and M. Alami, *Tetrahedron*, 2008, **64**, 4287-4294.
19. F. Albericio and J. Tulla-Puche, in *The Power of Functional Resins in Organic Synthesis*, Wiley-VCH Verlag GmbH & Co. KGaA, 2009, DOI: 10.1002/9783527626175.ch1, pp. 1-14.
20. V. R. Pattabiraman and J. W. Bode, *Nature*, 2011, **480**, 471-479.
21. A. Wilson, G. Gasparini and S. Matile, *Chem. Soc. Rev.*, 2014, **43**, 1948-1962.
22. H. M. Seifert, K. Ramirez Trejo and E. V. Anslyn, *J. Am. Chem. Soc.*, 2016, **138**, 10916-10924.
23. B. G. G. Lohmeijer and U. S. Schubert, *Macromol. Chem. Phys.*, 2003, **204**, 1072-1078.
24. R. Nishiyabu, Y. Kubo, T. D. James and J. S. Fossey, *Chem. Comm.*, 2011, **47**, 1106-1123.
25. F. Yakuphanoglu and B. F. Şenkal, *Polym. Eng. Sci.*, 2009, **49**, 722-726.